SINGULAR CASE OF PAIRS OF BILINEAR, QUADRATIC, OR HERMITIAN FORMS*

BY

L. E. DICKSON

The main object of this paper is a treatment of the equivalence of pairs of bilinear forms in the singular case by a purely rational method. The problem was first discussed by Kronecker,† who employed the irrational canonical form due to Weierstrass for the auxiliary non-singular case, instead of the rational canonical form (13) employed here. It is then a simple matter to deduce in Parts II and III the criteria for the equivalence of pairs of symmetric or Hermitian bilinear forms, or quadratic or Hermitian forms, in the singular case.

I. Pairs of bilinear forms in the singular case

Let ϕ and ψ be bilinear forms in the same variables $x_1, \cdots, x_r, y_1, \cdots, y_s$ with coefficients in a field F, such that ϕ and ψ are not equivalent in F to forms both of which involve fewer than r variables x_i or fewer than s variables y_i. We shall treat the singular case in which either $r \neq s$, or else $r=s$ and the determinant of $f=u\phi+v\psi$ is zero identically in u and v. If $r \neq s$, we take $r>s$ without loss of generality, since we may interchange the letters x and y.

Write f_i, ϕ_i, ψ_i for the partial derivatives of f, ϕ, ψ with respect to x_i. The linear functions $f_1, \cdots, f_r$ of $y_1, \cdots, y_s$ are linearly dependent. Hence there exist homogeneous polynomials A_i in u and v of the same degree with coefficients in F, such that the A_i are not all zero identically and such that $\sum A_i f_i \equiv 0$ identically in $u, v, y_1, \cdots, y_s$.

Lemma 1. *The degree of $A_1, \cdots, A_r$ in u and v is not changed if we apply to f non-singular linear transformations on the x's and the y's separately whose coefficients are independent of u and v.*

---

*Presented to the Society, December 26, 1924; received by the editors in July, 1926.

† Sitzungsberichte der Akademie der Wissenschaften, Berlin, 1890, pp. 1225–37. His brief paper was elaborated by Muth, *Elementartheiler*, 1899, pp. 93–133. Some minor errata in Muth's work were reproduced in the account in the Encyclopédie des Sciences Mathématiques, tome I, vol. II, 1911, pp. 412–25.

Evidently each A_i is unchanged if we transform the y's alone. Next, let f become g when the x's are expressed linearly in terms of new variables X_j. Then

$$X_j = \sum_{i=1}^{r} a_{ji}x_i, \quad f_i = \sum_{j=1}^{r} \frac{\partial g}{\partial X_j}\frac{\partial X_j}{\partial x_i} = \sum_{j=1}^{r} a_{ji}g_j,$$

and $\sum B_j g_j \equiv 0$, where $B_j = \sum_i a_{ji}A_i$ is of the same degree as the A_i.

Of all relations $\sum A_i f_i \equiv 0$, select one

$$\sum_{i=1}^{r} A_i f_i \equiv \sum_{\alpha=0}^{m} (-1)^{\alpha} C_\alpha u^\alpha v^{m-\alpha} \equiv 0 \tag{1}$$

in which the A_i are not all zero identically and are of minimal degree m in u and v, while

$$C_\alpha = \sum_{i=1}^{r} c_{\alpha i} f_i \qquad (\alpha = 0, 1, \cdots, m), \tag{2}$$

where the $c_{\alpha i}$ are numbers not all zero of the field F and are independent of u and v. By Lemma 1, m is an invariant of the pair ϕ and ψ under all non-singular linear transformations on the x's and the y's separately.

LEMMA 2. *$f_1, \cdots, f_r$ are linearly independent in F.*

For, if $\sum t_i f_i \equiv 0$, where the t_i are in F and are not all zero, the insertion of $f_i = u\phi_i + v\psi_i$ gives

$$\sum t_i \phi_i \equiv 0, \qquad \sum t_i \psi_i \equiv 0.$$

Let $t_e \neq 0$ and apply the non-singular transformation

$$x_e = t_e X_e, \quad x_j = X_j + t_j X_e \quad (j = 1, \cdots, r; j \neq e).$$

Then $\phi = \sum x_i \phi_i$ becomes

$$\sum_{j \neq e} X_j \phi_j + X_e \sum_{i=1}^{r} t_i \phi_i,$$

which lacks X_e since the final sum is zero. Likewise, ψ lacks X_e. But this contradicts our initial assumption.

COROLLARY. *The invariant m exceeds zero.*

For, if $m = 0$, (1) becomes $C_0 = 0$ and (2) implies that every $c_{\alpha i}$ is zero, contrary to hypothesis.

LEMMA 3. *$C_0, \cdots, C_m$ are linearly independent in F.*

Suppose that $a_0, \cdots, a_m$ are numbers, not all zero, of F such that

$$\sum_{\alpha=0}^{m} a_\alpha C_\alpha \equiv 0. \tag{3}$$

Denote the first and second sums in (1) by Q and R. Write

$$P = \sum_{\alpha=0}^{m} (-1)^\alpha a_\alpha u^{m-\alpha} v^\alpha. \tag{4}$$

Each PA_i is a polynomial in u and v of degree $2m$; let k_i be the coefficient of $u^m v^m$. The terms $\sum k_i u^m v^m f_i$ of PQ are obtained from the terms $\sum a_\alpha C_\alpha u^m v^m$ of PR by replacing C_α by its value (2). Hence (3) implies $\sum k_i f_i \equiv 0$. By Lemma 2, each $k_i = 0$. Hence $PQ \equiv 0$ may be written in the form

$$v^{m+1} \sum f_i g_i + u^{m+1} \sum f_i h_i \equiv 0 \qquad (i = 1, \cdots, r), \tag{5}$$

in which the g_i and h_i are polynomials of degree $m-1$ in u and v. If they were all zero identically, every PA_i and hence every A_i would be zero identically. Replacing f_i by its value $u\phi_i + v\psi_i$, we see that the first sum in (5) is of degree $\leqq m$ in u, while the second sum is evidently of degree $> m$ in u. Hence each sum is zero identically, and one of them gives a linear relation between $f_1, \cdots, f_r$ the degree in u and v of whose coefficients is $m-1$, while those coefficients (g_i or h_i) are not all zero identically. But this contradicts the definition of m as the minimal degree.

By Lemmas 2 and 3, the matrix of the coefficients in (2) has a non-vanishing determinant of order $m+1$. Hence we can annex further rows, also composed of numbers of F, and obtain a square matrix

$$(c_{\alpha\beta}) \qquad (\alpha = 0, 1, \cdots, r-1\,;\, \beta = 1, \cdots, r)$$

whose determinant is not zero. Let the non-singular transformation

$$x_\beta = \sum_{\alpha=0}^{r-1} c_{\alpha\beta} X_\alpha \qquad (\beta = 1, \cdots, r), \tag{6}$$

whose coefficients are independent of u and v, replace f, ϕ, ψ by F, Φ, Ψ. The partial derivatives of the latter with respect to X_i are denoted by F_i, Φ_i, Ψ_i. Then

$$F_\alpha = \sum_{\beta=1}^{r} \frac{\partial f}{\partial x_\beta} \frac{\partial x_\beta}{\partial X_\alpha} = \sum_{\beta=1}^{r} c_{\alpha\beta} f_\beta = C_\alpha, \tag{7}$$

by (2). By (1),

$$\sum (-1)^\alpha F_\alpha u^\alpha v^{m-\alpha} \equiv 0.$$

Replacing F_α by $u\Phi_\alpha+v\Psi_\alpha$, and equating to zero the coefficients of v^{m+1}, uv^m, u^2v^{m-1}, $\cdots$, u^mv, u^{m+1}, we get

$$\Psi_0 \equiv 0,\ \Psi_1 \equiv \Phi_0,\ \Psi_2 \equiv \Phi_1,\ \cdots,\ \Psi_m \equiv \Phi_{m-1},\ \Phi_m \equiv 0. \tag{8}$$

Hence

$$F_0 \equiv u\Psi_1,\ F_1 \equiv u\Psi_2 + v\Psi_1,\ \cdots,\ F_{m-1} \equiv u\Psi_m + v\Psi_{m-1},\ F_m \equiv v\Psi_m.$$

Multiply the first k equations by v^{k-1}, $-uv^{k-2}$, $\cdots$, $(-u)^{k-2}v$, $(-u)^{k-1}$, respectively, and add; we get

$$-(-u)^k\Psi_k \equiv \sum_{i=0}^{k-1} F_i(-u)^i v^{k-i-1} \qquad (k = 1, \cdots, m). \tag{9}$$

If $\Psi_1, \cdots, \Psi_m$ satisfy a linear relation whose coefficients are in the field and are not all zero, multiply it by u^m and apply (9). We get a linear relation between $F_0, \cdots, F_{m-1}$ whose coefficients are polynomials, not all identically zero, of degree $m-1$ in u and v. Inserting the values (7) and applying (2), we get a similar linear relation between $f_1, \cdots, f_r$, contrary to the definition of m. Hence $\Psi_1, \cdots, \Psi_m$ are linearly independent in the field. We may therefore choose linear functions $Y_{m+1}, \cdots, Y_s$ of $y_1, \cdots, y_s$ with coefficients in the field F such that the Ψ and Y are linearly independent in F. Then

$$\eta_\alpha = \Psi_\alpha \qquad (\alpha = 1, \cdots, m), \quad \eta_t = Y_t \qquad (t = m+1, \cdots, s) \tag{10}$$

define a non-singular linear transformation on $y_1, \cdots, y_s$ with coefficients in F. Transformation (6) replaces $u\phi+v\psi$ by

$$u\Phi + v\Psi = \sum_{i=0}^{r-1} u\Phi_i X_i + \sum_{i=0}^{r-1} v\Psi_i X_i.$$

For $i=0, \cdots, m$, change i to $\alpha-1$ in the first sum, and i to α in the second. For $i>m$, change i to p. Then by (8) and (10),

$$u\Phi + v\Psi = \sum_{\alpha=1}^{m} (uX_{\alpha-1} + vX_\alpha)\eta_\alpha + \sum_{p=m+1}^{r-1} (u\Phi_p + v\Psi_p)X_p.$$

Here Φ_p and Ψ_p are linear in $y_1, \cdots, y_s$ and hence in $\eta_1, \cdots, \eta_s$ by (10):

$$\Phi_p = \sum_{\alpha=1}^{s} a_{p\alpha}\eta_\alpha, \quad \Psi_p = \sum_{\alpha=1}^{s} b_{p\alpha}\eta_\alpha \quad (p = m+1, \cdots, r-1),$$

where the a's and b's are numbers in F. We have

$$
\begin{aligned}
u\Phi + v\Psi = {} & u\Big(X_0 + \sum_p a_{p1}X_p \Big)\eta_1 + u\sum_{\alpha=2}^{m} X_{\alpha-1}\eta_\alpha \\
& + u\sum_{\alpha=2}^{m}\sum_p b_{p\,\alpha-1}X_p\eta_\alpha + v\sum_{\alpha=1}^{m}\Big(X_\alpha + \sum_p b_{p\alpha}X_p\Big)\eta_\alpha + \rho + \sigma,
\end{aligned} \tag{11}
$$

where the summations with respect to p are from $m+1$ to $r-1$, and $\rho = u(l_2\eta_2 + \cdots + l_m\eta_m)$, the l's being linear in the X_p $(p>m)$, while σ is a bilinear form in the X_p, η_q $(p>m,\ q>m)$. In deriving (11), the terms involving η_α $(\alpha>m)$ which arose from Ψ_p were carried to σ; the terms involving $\eta_\alpha(\alpha=2, \cdots, m)$ which arose from Φ_p were carried to ρ, while those involving $\eta_\alpha(\alpha>m)$ were carried to σ. Finally we added the terms in the third sum in the second member and subtracted the same from ρ. To (11) apply the non-singular transformation

$$
\xi_0 = X_0 + \sum_p a_{p1}X_p, \quad \xi_\alpha = X_\alpha + \sum_p b_{p\alpha}X_p, \quad \xi_p = X_p
$$
$$
(\alpha = 1, \cdots, m\,;\ p = m+1, \cdots, r-1).
$$

We get (12) in which $u\Phi+v\Psi$ is written for σ.

THEOREM 1. *If ϕ and ψ are bilinear forms in $x_1, \cdots, x_r, y_1, \cdots, y_s$ with coefficients in a field F, and $r \geqq s$, and when $r=s$ the determinant $|u\phi+v\psi|$ is zero identically in u and v, there exist non-singular linear transformations on the x's and on the y's separately with coefficients in F which reduce $u\phi+v\psi$ to*

$$
S = \sum_{\alpha=1}^{m}(u\xi_{\alpha-1} + v\xi_\alpha)\eta_\alpha + u\sum_{\alpha=2}^{m} l_\alpha\eta_\alpha + u\Phi + v\Psi, \tag{12}
$$

where $l_2, \cdots, l_m$ are linear forms in $\xi_{m+1}, \cdots, \xi_{r-1}$, and Φ and Ψ are bilinear forms in $\xi_{m+1}, \cdots, \xi_{r-1}, \eta_{m+1}, \cdots, \eta_s$, all with coefficients in F.

If $r<s$, $f=u\phi+v\psi$ can be reduced to a form derived from (12) by interchanging the letters ξ and η, since the form derived from f by interchanging the letters x and y can be reduced to (12).

Given a pair of bilinear forms λ and μ with coefficients in F such that $u\lambda+v\mu$ is of rank R, we can select independent linear combinations ϕ and ψ of λ and μ with coefficients in F such that the greatest common divisor of all R-rowed determinants of $u\phi+v\psi$ is not divisible by u, whence ψ is of rank R. In the corresponding form (12), let τ be the rank of $u\Phi+v\Psi$.

LEMMA 4. *The greatest common divisor of all τ-rowed determinants of $u\Phi+v\Psi$ is not divisible by u.*

In the matrix M of (12), the first m columns form the matrix M_1 of the coefficients of $\xi_0, \cdots, \xi_{r-1}$ in

$$\frac{\partial S}{\partial \eta_1} = u\xi_0 + v\xi_1, \quad \frac{\partial S}{\partial \eta_2} = u\xi_1 + v\xi_2 + ul_2, \quad \cdots, \quad \frac{\partial S}{\partial \eta_m} = u\xi_{m-1} + v\xi_m + ul_m.$$

Since these are evidently linearly independent, M_1 is of rank m. The last $s-m$ columns of M form the matrix M_2 of the coefficients of $\xi_0, \cdots, \xi_{r-1}$ in the partial derivatives of $u\Phi+v\Psi$ with respect to $\eta_{m+1}, \cdots, \eta_s$. Hence by the definition of τ, the rank of M_2 is τ. Thus not every $(m+\tau)$-rowed determinant of M is zero. But every determinant with more than $m+\tau$ columns is zero, since it contains more than τ columns of M_2. Hence the rank of M is $m+\tau=R$. Every non-vanishing $(m+\tau)$-rowed determinant of M contains exactly τ columns of M_2 and is a linear function of τ-rowed determinants of M_2 and hence of $u\Phi+v\Psi$. If all τ-rowed determinants of the latter were divisible by u, the same would be true of all determinants with $m+\tau=R$ columns of M and hence of S, contrary to the remark before Lemma 4.

First, let $u\Phi+v\Psi$ fall under the non-singular case, so that it has the same number of variables in the two series and its determinant is not zero identically in u and v. By Lemma 4, this determinant is not divisible by u, whence $|\Psi| \neq 0$. Then there exist non-singular transformations on the variables of the two series separately, with coefficients in F, which reduce Φ and Ψ to the respective forms*

$$(13) \qquad \rho = \sum_k \left(\sum_{i=1}^{a_k-1} x_{ki}y_{k\,i+1} + x_{ka_k} \sum_{i=1}^{a_k} b_{ki}y_{ki} \right), \quad \sigma = \sum_k \sum_{i=1}^{a_k} x_{ki}y_{ki}.$$

The transformations employed leave unaltered each variable occurring in the first sum in (12) and replace the second sum by $u\sum t_\alpha\eta_\alpha$, where t_α is a linear function f the x_{ki}. Hence if $r \geqq s$, ϕ and ψ may be reduced rationally to

$$(14) \qquad J = \sum_{\alpha=1}^{m} \xi_{\alpha-1}\eta_\alpha + \sum_{\alpha=2}^{m} \sum_k \sum_{i=1}^{a_k} d_{kia}x_{ki}\eta_\alpha + \rho, \quad K = \sum_{\alpha=1}^{m} \xi_\alpha\eta_\alpha + \sigma.$$

Second, let $u\Phi+v\Psi$ fall under the singular case. We apply to it our initial discussion of $u\phi+v\psi$ and conclude that it can be transformed ration-

*Dickson, *Modern Algebraic Theories*, Chicago, Sanborn, 1926, p. 116, Theorem IV. A direct verification follows from the fact that the invariant factors other than 1 of $\rho-\lambda\sigma$ are

$$\lambda^{a_k} - \sum_{i=1}^{a_k} b_{ki}\lambda^{i-1} \qquad (k = 1, 2, \cdots),$$

which may be identified with any assigned polynomials.

ally into a form $R_1+uP_1+f_1$ of type (12) or a form $S_1+uQ_1+f_1$ obtained from it by interchanging the letters ξ and η, where

$$f_1 = u\Phi_1 + v\Psi_1, \quad R_1 = \sum_{\alpha=1}^{m_1}(u\xi_{1\,\alpha-1} + v\xi_{1\alpha})\eta_{1\alpha}, \quad P_1 = \sum_{\alpha=2}^{m_1} P_{1\alpha}\eta_{1\alpha},$$

$$S_1 = \sum_{\alpha=1}^{n_1}(u\eta_{1\,\alpha-1} + v\eta_{1\alpha})\xi_{1\alpha}, \quad Q_1 = \sum_{\alpha=2}^{n_1} q_{1\alpha}\xi_{1\alpha}.$$

If f_1 falls under the non-singular case, Φ_1 and Ψ_1 can be transformed rationally into forms of type (13). If f_1 falls under the singular case, another repetition of the initial discussion shows that f_1 can be transformed rationally into $R_2+uP_2+f_2$ or $S_2+uQ_2+f_2$, where $f_2, R_2, \cdots, Q_2$ are of the type $f_1, R_1, \cdots, Q_1$ with the first subscripts 1 changed to 2.

This process must terminate, since the numbers of variables in $\Phi, \Phi_1, \cdots$ form a series of decreasing positive integers.

To prove by induction that $\sum l_\alpha\eta_\alpha$ (in (12)), P_1 or Q_1, P_2 or Q_2, etc., can be removed by a linear transformation with coefficients in F which does not alter the sum of the remaining terms, we assume that all but $\sum l_\alpha\eta_\alpha$ have been so removed and shall prove that we can then remove it also.

We first show that $R_1, S_1, R_2, \cdots$ can all be written as terms of $u\rho+v\sigma$ in (13) by employing new values of k. To do this for R_1, write

$$\xi_{1\alpha} = x_{k\,\alpha+1}\ (\alpha = 0,1,\cdots,m_1), \quad \eta_{1\alpha} = y_{k\,\alpha+1} \quad (\alpha = 1,\cdots,m_1),$$

and $m_1=a_k-1$. Then R_1 becomes

$$u\sum_{i=1}^{a_k-1} x_{ki}y_{k\,i+1} + v\sum_{i=1}^{a_k} x_{ki}y_{ki} \tag{15}$$

for $y_{k1}=0$; the coefficients of u and v are the terms of (13) for the present value of k and with every $b_{ki}=0$. In S_1 write

$$\xi_{1\alpha} = x_{k\,n_1+1-\alpha}\ (\alpha = 1,\cdots,n_1), \quad \eta_{1\alpha} = y_{k\,n_1+1-\alpha} \quad (\alpha = 0,1,\cdots,n_1),$$

and $n_1=a_k-1$. Then S_1 becomes (15) for $x_{ka_k}=0$.

Whether $u\Phi+v\Psi$ falls under the singular or non-singular case, we may therefore reduce the initial pair ϕ, ψ with $r \geqq s$ to a pair (14) such that, for certain values of k, either x_{ka_k} is suppressed or y_{k1} is suppressed and every b_{ki} replaced by zero in (13) and the second sum in (14).

It remains to prove that we may take every d to be zero in (14). We shall first prove that we may replace $d_{ki\alpha}$ by zero when $i<a_k$. We proceed by induction on α, keeping k fixed. For a given integer n such that $2 \leqq n \leqq m$,

assume that $d_{kia}=0$ for $i<a_k$, $\alpha=2, \cdots, n-1$. Apply the transformation*

$$\xi_n = X_n + \sum_{i=2}^{a_k} t_{ki}x_{ki}, \qquad y_{ki} = Y_{ki} - t_{ki}\eta_n \qquad (i = 2, \cdots, a_k)$$

which leaves unaltered the variables not occurring in the left members. The increment to K is zero, while that to J is

$$\eta_{n+1}\sum_{i=2}^{a_k} t_{ki}x_{ki} - \sum_{i=1}^{a_k-1} t_{k\,i+1}x_{ki}\eta_n + cx_{ka_k}\eta_n,$$

where the first sum is to be omitted if $n=m$. Taking

$$t_{k\,i+1} = d_{kin} \qquad (i = 1, \cdots, a_k - 1),$$

we obtain forms of type (14) in which $d_{kia}=0$ for $i<a_k$, $\alpha=2, \cdots, n$. Hence the induction is complete and we now have

$$(16) \qquad J = \sum_{\alpha=1}^{m}\xi_{\alpha-1}\eta_\alpha + \sum_{\alpha=2}^{m}\sum_{k} e_{k\alpha}x_{ka_k}\eta_\alpha + \rho, \; K = \sum_{\alpha=1}^{m}\xi_\alpha\eta_\alpha + \sigma.$$

We shall prove that we may replace each $e_{k\alpha}$ by zero. No proof is needed when k is such that x_{ka_k} is to be suppressed. We proceed by induction on α, keeping k fixed. For a given positive integer n $(n<m)$, assume that $e_{k\alpha}=0$ for $\alpha=m, m-1, \cdots, n+2$.

First, let $n<a_k$ and apply the transformation

$$(17) \qquad \begin{aligned} \xi_{n-j} &= X_{n-j} + t\,x_{ka_k-j} \qquad (j = 0, \cdots, n), \\ y_{ka_k-j} &= Y_{ka_k-j} - t\eta_{n-j} \qquad (j = 0, \cdots, n-1), \end{aligned}$$

which leaves unaltered the remaining variables. The increment to K is zero, while that to J cancels to

$$(18) \qquad tx_{ka_k}\eta_{n+1} - tx_{ka_k}\sum_{j=0}^{n-1} b_{ka_k-j}\eta_{n-j}.$$

Taking $t=-e_{k\,n+1}$, we obtain (16) with $e_{k\alpha}=0$ $(\alpha=m, \cdots, n+1)$, but with a new term $cx_{ka_k}\eta_1$. The latter is removed by the transformation $\xi_0 = \Xi_0 - cx_{ka_k}$.

Second, let k be one of the values for which y_{k1} is not to be suppressed from (13) and let $n\geqq a_k$. Apply transformation (17) with $j=0, \cdots, a_k-1$ in both parts. Then K is unaltered, while the increment to J cancels to

---

*But let the summation extend only to a_k-1 if k is such that x_{ka_k} is to be suppressed from (13) and (14). Then the text holds with $c=0$.

(18) in which now j takes the values $0, \cdots, a_k-1$. We proceed as in the first case.

In either case we obtain forms of type (16) in which now $e_{k\alpha}=0$ ($\alpha = m, m-1, \cdots, n+1$). By induction on n, we obtain (16) in which every $e_{k\alpha}=0$.

Third, let k be a value for which y_{k1} is to be suppressed from (16), while every $b_{ki}=0$. Then $m<a_k$, whence $n<a_k$ and the second case is excluded. In the proof, we write

$$S = uJ + vK, \Sigma_\alpha = \frac{\partial S}{\partial \xi_\alpha} \quad (\alpha = 0, \cdots, m), \quad S_{ki} = \frac{\partial S}{\partial x_{ki}} \quad (i = 1, \cdots, a_k).$$

By (16) and (13),

$$(19) \quad \Sigma_0 = u\eta_1, \quad \Sigma_\alpha = u\eta_{\alpha+1} + v\eta_\alpha \quad (\alpha = 1, \cdots, m-1), \quad \Sigma_m = v\eta_m,$$

the solution of which is

$$(20) \qquad u^{\alpha+1}\eta_{\alpha+1} = \sum_{j=0}^{\alpha} (-1)^j u^{\alpha-j} v^j \Sigma_{\alpha-j} \qquad (\alpha = 0, \cdots, m-1).$$

Eliminating the resulting value of η_m from the last equation (19), we get

$$(21) \qquad u^m \Sigma_m = \sum_{i=1}^{m} (-1)^{i-1} u^{m-i} v^i \Sigma_{m-i}.$$

Next,

$$(22) \qquad S_{ki} = uy_{k\,i+1} + vy_{ki}\,(i < a_k), \; S_{ka_k} = u \sum_{\alpha=2}^{m} e_{k\alpha}\eta_\alpha + vy_{ka_k}.$$

The former with y_{k1} suppressed gives

$$(23) \qquad u^{i+1} y_{k\,i+2} = \sum_{j=0}^{i} (-1)^j u^{i-j} v^j S_{k\,i+1-j} \quad (i = 0, \cdots, a_k - 2).$$

By (20) and (23), $u^t\eta_t$ and $u^t y_{k\,t+1}$ are linear functions of the Σ's and S's whose coefficients are polynomials in u and v of degree $t-1$. Multiply the last equation (22) by u^{m-1}. If $m \geqq a_k$, we get

$$(24) \qquad u^{m-1} S_{ka_k} = \sum_{\alpha=2}^{m} e_{k\alpha} u^{m-\alpha} u^\alpha \eta_\alpha + vu^{m-a_k} u^{a_k-1} y_{ka_k}.$$

By the remark just made, this yields a linear relation between the first derivatives of S whose coefficients are polynomials in u and v of degree $m-1$. But the S on the left does not occur on the right. This contradicts the definition of m.

This completes the proof that we may take every d to be zero in (14). In stating the following theorem, we keep separate the forms R_i and S_i previously combined with $u\rho+v\sigma$. For each i only one of the pair R_i and S_i is a component of the canonical form. We shall change the notation of the variables occurring in a component S_j.

THEOREM 2. *Let ϕ and ψ be bilinear forms with coefficients in a field F such that ψ has* the same rank as $f=u\phi+v\psi$. In the singular case, f can be reduced by non-singular linear transformations with coefficients in F to $S=\sum R_i+\sum S_j+u\rho+v\sigma$, where ρ and σ are given by* (13), *and*

$$(25) \qquad R_i = \sum_{\alpha=1}^{m_i}(u\xi_{i\,\alpha-1}+v\xi_{i\alpha})\eta_{i\alpha}, \quad S_j=\sum_{\alpha=1}^{n_j}(uY_{j\,\alpha-1}+vY_{j\alpha})X_{j\alpha}.$$

No variable appears in two of the forms $R_1, R_2, \cdots, S_1, S_2, \cdots, u\rho+v\sigma$, while all their variables are linearly independent in F.

We shall now prove that the invariant factors other than 1 of the matrix M of S coincide with those of $u\rho+v\sigma$. The determinant D of order d of the latter is not zero identically in u and v. Write $t=\sum m_i+\sum n_j$. Then the rank of M will be shown to be $d+t$. In M, the coefficient of a product $\xi\eta$ is placed at the intersection of the row $[\xi]$ corresponding to ξ and the column $[\eta]$ corresponding to η. Similarly for XY.

First, delete the rows $[\xi_{im_i}]$ corresponding to the ξ's with maximal second subscripts in the R_i, and the columns $[Y_{jn_j}]$ corresponding to Y's with maximal second subscripts in the S_j. We obtain a square matrix which (apart from elements of D) has every diagonal element equal to u, and has zeros as elements to the right (left) of these u's in the case of $R_i(S_j)$, whence its determinant is u^tD.

Second, delete the rows $[\xi_{i0}]$ and the columns $[Y_{j0}]$. We obtain a square matrix which (apart from elements of D) has every diagonal element equal to v, and has zeros as elements to the left (right) of these v's in the case of $R_i(S_j)$, whence its determinant is v^tD.

Hence the greatest common divisor of all $(d+t)$-rowed determinants of M is D. This yields†

LEMMA 5. *The invariant factors other than unity of S coincide with those of $u\rho+v\sigma$.*

Hence ρ and σ are uniquely determined by the invariant factors of $u\phi+v\psi$.

*We saw that this property may be secured by passing to suitably chosen linear combinations of ϕ and ψ with coefficients in F.

†See the writer's *Modern Algebraic Theories*, pp. 96, 97.

The variables ξ, X, x of the first series in S are linearly independent; likewise for the variables η, Y, y of the second series (Theorem 2). Write $W=u\rho+v\sigma$. No variable is common to two of the R_i, S_j, W. Hence a linear function $[S]$ of the first derivatives of $S=W+\sum R_i+\sum S_j$ with respect to the variables of either series is a sum of linear functions $[W]$, $[R_1]$, $[R_2]$, $\cdots$, $[S_1]$, $\cdots$ of the first derivatives of W, R_1, R_2, $\cdots$ with respect to their own variables only. Also, $[S]=0$ implies $[W]=0$, $[R_1]=0$, $\cdots$. Since $|W|$ is not identically zero, the vanishing of $[W]$ implies that its coefficients are all zero.

If we suppress the first subscript i from m_i and the variables in R_i, and write Σ_α for $\partial R_i/\partial\xi_\alpha$, we have (19)-(21). Since Σ_0, $\cdots$, Σ_{m-1} are linearly independent, every linear relation between the Σ's is obtained by multiplying (21) by a function of u and v.

The first derivatives of R_i with respect to the $\eta_{i\alpha}$ are evidently linearly independent. Since R_i becomes S_i by replacing ξ, η, m by Y, X, n, we have

LEMMA 6. *As a complete set of independent linear relations between the first derivatives of S with respect to the variables ξ, X, x of the first series we may take*

$$u^{m_i}\frac{\partial S}{\partial \xi_{im_i}}=\sum_{j=1}^{m_i}(-1)^{j-1}u^{m_i-j}v^j\frac{\partial S}{\partial \xi_{i\,m_i-j}}, \tag{26}$$

where i takes the same values as the subscript of R_i. Here m_i is the minimal degree in u and v of the coefficients in any linear relation between the first derivatives of S with respect to the variables ξ_i of R_i. Similarly, n_j is the minimal degree in u and v of the coefficients in any linear relation between the first derivatives of S with respect to the variables Y_j of S_j and these relations of minimal degrees form a complete set of independent linear relations between the first derivatives of S with respect to the variables η, Y, y of the second series.

These minimal degrees m_i and n_j uniquely determine the R_i and S_j. By Lemma 1, they have the same value for $u\phi+v\psi$ as for the canonical form S. They are called the *minimal numbers* of $u\phi+v\psi$ or of S.

By Lemmas 5 and 6, and the remarks just following them, we conclude that two pairs of bilinear forms ϕ, ψ and λ, μ in F, each satisfying the initial assumption and having the property about rank in Theorem 2, have the same canonical form and hence are equivalent in F if and only if they have the same minimal numbers and the same invariant factors.

Our next step is to prove this result when the two restrictions on the forms are removed.

Let f and g be any bilinear forms with coefficients in F in the singular case. After supplying zero coefficients, if necessary, we may assume that

there is the same number n of variables X_i as variables Y_i. If τ is the rank of $h=uf+vg$, there are exactly $n-\tau$ independent linear relations between the first derivatives of h with respect to $X_1, \cdots, X_n$ (or $Y_1, \cdots, Y_n$). Let them be chosen so that their coefficients are of the least possible degrees $M_1, \cdots, M_{n-\tau}$ (or $N_1, \cdots, N_{n-\tau}$) in u and v. Let the M's be arranged in order of increasing magnitude; likewise the N's. The resulting numbers M_i and N_i are called the *minimal numbers* of the pair f, g.

For example, let $f=X_1Y_3+X_3Y_1$, $g=X_2Y_3+X_3Y_2$. Then h has the matrix

$$\begin{pmatrix} 0 & 0 & u \\ 0 & 0 & v \\ u & v & 0 \end{pmatrix}$$

of rank 2. Hence $vh_1-uh_2\equiv 0$ is the single linear relation between the first partial derivatives h_1, h_2, h_3, of h with respect to X_1, X_2, X_3. The minimal numbers are $M_1=1$, $N_1=1$. Since the two-rowed determinants include $-u^2$ and $-v^2$, their greatest common divisor is 1. Hence the matrix has no invariant factor other than 1.

If the first ρ of the M's and the first σ of the N's are zero, while the others are positive, we can transform f, g into a pair ϕ, ψ involving only $r=n-\rho$ variables x_i and only $s=n-\sigma$ variables y_i (Lemma 2). Then ϕ, ψ satisfy our initial assumption that they cannot be transformed rationally into a pair of forms both of which involve fewer than r variables x_i or fewer than s variables y_i.

Since the M_i and N_i are invariant under every linear transformation on the X's or on the Y's (Lemma 1), the positive ones of M_i and N_i coincide with the above m_i and n_i of ϕ and ψ.

Introduce the linear combinations

$$\phi' = a\phi + b\psi, \quad \psi' = c\phi + d\psi \qquad (ad - bc = 1). \tag{27}$$

Write $\chi=u\phi+v\psi$, $\chi'=u'\phi'+v'\psi'$. Then $\chi=\chi'$ if

$$u' = du - cv, \quad v' = -bu + av. \tag{28}$$

Before Lemma 4, we chose a, b, c, d so that the greatest common divisor of all R-rowed determinants of χ' is not divisible by u', where R is the rank of χ and hence of ψ'.

Let λ, μ be another pair of bilinear forms in $x_1, \cdots, x_r, y_1, \cdots, y_s$ with coefficients in F, having the same minimal numbers and same invariant factors as the pair ϕ, ψ. Using the same a, b, c, d as in (27), write

$$\lambda' = a\lambda + b\mu, \quad \mu' = c\lambda + d\mu, \quad \nu = u\lambda + v\mu, \quad \nu' = u'\lambda' + v'\mu'.$$

Since χ' and ν' are transformed into χ and ν by (28), they have the same minimal numbers and same invariant factors, and in particular the same rank R and the same greatest common divisor of their R-rowed determinants. Hence also μ' is of rank R.

Theorem 2 and the remarks following Lemmas 5 and 6 show that χ' and ν' have the same canonical form and hence are equivalent in F. Thus the pairs ϕ, ψ and λ, μ are equivalent in F. This completes the proof of the fundamental

THEOREM 3. *In the singular case, two pairs of bilinear forms in the same $n+n$ variables with coefficients in any field F are equivalent in F if and only if they have the same invariant factors and the same minimal numbers.*

The canonical pair shows that these may be given any assigned values such that each invariant factor divides the next.

II. PAIRS OF SYMMETRIC OR HERMITIAN BILINEAR FORMS IN THE SINGULAR CASE

In order to treat simultaneously symmetric and Hermitian bilinear forms, we employ $\bar{a}$ to denote a or the conjugate imaginary to a in the respective cases. Consider the form

$$f = \sum_{i,j=1}^{n} a_{ij}x_iy_j, \qquad a_{ji} = \bar{a}_{ij}, \tag{29}$$

whose matrix $A=(a_{ij})$ has the property $A'=\tilde{A}$. Then

$$\frac{\partial f}{\partial x_i} = \sum_{j=1}^{n} a_{ij}y_j, \quad \frac{\partial f}{\partial y_i} = \sum a_{ji}x_j = \sum_{j=1}^{n} \bar{a}_{ij}x_j.$$

If

$$\sum_{i=1}^{n} c_i \frac{\partial f}{\partial x_i} \equiv 0, \tag{30}$$

where the c_i are independent of the x's and y's, then

$$\sum_{i=1}^{n} c_i a_{ij} = 0 \qquad (j = 1, \cdots, n), \tag{31}$$

whence $\sum \bar{c}_i \, \partial f/\partial y_i \equiv 0$. Take $f=u\phi+v\psi$. Hence each minimal number M_i of ϕ, ψ coincides with a minimal number N_i and vice versa.

Consider two pairs of forms ϕ, ψ and λ, μ whose n-rowed matrices A, B, C, D have the property $M'=\tilde{M}$. Let $u\phi+v\psi$ and $u\lambda+v\mu$ have the same minimal numbers M_i (and hence the same N_i) and the same invariant factors. Then the pairs are equivalent by Theorem 3, so that there exist non-singular n-rowed matrices P and Q for which

$$C=PAQ, \quad D=PBQ.$$

Hence* there exists a non-singular n-rowed matrix E such that

$$C = E'A\tilde{E}, \quad D = E'B\tilde{E},$$

whence the linear transformation on the x's with the matrix E and that on the y's with the matrix $\tilde{E}$ replace ϕ by λ and ψ by μ. Since this argument may be reversed, we have

THEOREM 4. *In the singular case, two pairs of symmetric (or Hermitian) bilinear forms are equivalent under a non-singular linear transformation on the x's and the cogredient (or conjugate) transformation on the y's if and only if they have the same invariant factors and the same minimal numbers M_i (which here coincide with the N_i).*

In seeking a canonical pair for such a pair ϕ, ψ, note that the number of the M_i which have the value zero is the number of the x's (or y's) which can be removed by linear transformation. Each M_i which exceeds zero will be denoted by m_i; let their number be p. In the notations of Theorem 2, a canonical form is

$$C = \sum_{i=1}^{p} T_i + u\gamma + v\delta, \quad T_i = R_i + S_i,$$

with $n_i = m_i$ in S_i, while $u\gamma + v\delta$ is symmetric (or Hermitian), has a determinant not zero identically in u and v, and has the same invariant factors as $u\phi + v\psi$. Then the latter is equivalent to C in the sense of Theorem 4.

To emphasize the fact that each T_i is symmetric (or Hermitian), write x_α for ξ_α, $x_{m_i+\alpha}$ for X_α, y_α for Y_α, and $y_{m_i+\alpha}$ for η_α.

THEOREM 5. *In the singular case, any pair of symmetric (or Hermitian) bilinear forms can be reduced by a non-singular linear transformation on the x's and the cogredient (or conjugate) transformation on the y's to a canonical pair which are the coefficients of u and v in $\sum T_i + u\gamma + v\delta$, where*

$$T_i = \sum_{\alpha=1}^{m_i} [(ux_{i\,\alpha-1} + vx_{i\alpha})y_{i\,m_i+\alpha} + (uy_{i\,\alpha-1} + vy_{i\alpha})x_{i\,m_i+\alpha}]$$

is symmetric and Hermitian in its $2m_i+1$ pairs of variables, while γ, δ is a canonical pair† for the non-singular case.

Modern Algebraic Theories, pp. 122–3.

†For the quadratic case, we may employ Weierstrass's canonical pair, *Modern Algebraic Theories*, p. 131. For the Hermitian case, we may employ the corresponding pair, Logsdon, American Journal of Mathematics, vol. 44 (1922), p. 254. Or, without separating the two cases we may employ the writer's pair, *Modern Algebraic Theories*, pp. 126–8. In all of these cases, γ and δ have any prescribed invariant factors (or elementary divisors).

III. Pairs of quadratic or Hermitian forms in the singular case

Let $f=u\phi+v\psi$ be of type (29). To it corresponds

$$h = \sum_{i,j=1}^{n} a_{ij}x_i\tilde{x}_j, \quad a_{ji} = \tilde{a}_{ij}, \tag{32}$$

which is a quadratic or Hermitian form according as $\tilde{a}$ is identified with a or its conjugate imaginary, and likewise for $\tilde{x}$. In the respective cases, write $g=2$ or 1. Then

$$\frac{\partial h}{\partial x_i} = g\sum_{j=1}^{n} a_{ij}\tilde{x}_j, \quad \sum_{i=1}^{n} c_i\frac{\partial h}{\partial x_i} = g\sum_{j=1}^{n}\left(\sum_{i=1}^{n} c_i a_{ij}\right)\tilde{x}_j = 0,$$

by (31). Hence any linear relation (30) between the first derivatives of f with respect to $x_1, \cdots, x_n$, whose coefficients c_i are polynomials in u and v, implies a linear relation with the same coefficients c_i between the first derivatives of h, and conversely. Hence the minimal numbers $M_i=N_i$ of f coincide with those of h. By* Theorem 4, we have

Theorem 6. *In the singular case, two pairs of quadratic or Hermitian forms are equivalent if and only if they have the same invariant factors and the same minimal numbers M_i.*

From Theorem 5 we obtain a canonical pair of Hermitian forms by identifying each y with the conjugate imaginary to x, and a canonical pair of quadratic forms by identifying each y with x. The latter pair is

$$\gamma + 2\sum_{i=1}^{p}\sum_{\alpha=1}^{m_i} x_{i\,\alpha-1}x_{i\,m_i+\alpha}, \quad \delta + 2\sum_{i=1}^{p}\sum_{\alpha=1}^{m_i} x_{i\alpha}x_{i\,m_i+\alpha}, \tag{33}$$

where γ, δ is the canonical pair in the non-singular case (last foot-note in II). In particular, there exists a pair of quadratic or Hermitian forms with any prescribed invariant factors and any prescribed minimal numbers.

An example of (33) is given by $2x_0x_2$, $2x_1x_2$. We have

$$v\frac{\partial f}{\partial x_0} - u\frac{\partial f}{\partial x_1} \equiv 0, \quad f = 2ux_0x_2 + 2vx_1x_2.$$

Here there is a single minimal number $M_1=1$ and no invariant factor.

---

*And the corollaries on pp. 66 and 68 of *Modern Algebraic Theories*.

University of Chicago,
Chicago, Ill.

TRIADS OF RULED SURFACES*

BY

A. F. CARPENTER

The projective differential geometry of a configuration composed of two ruled surfaces whose generators are in one-to-one correspondence has been discussed by E. P. Lane†. He uses for defining system of equations a set of four ordinary linear first order differential equations in four dependent variables which, with slightly changed notation, we will write in the form

$$
\begin{aligned}
y' &= a_{11}y + a_{12}z + a_{13}\alpha + a_{14}\beta, \\
z' &= a_{21}y + a_{22}z + a_{23}\alpha + a_{24}\beta, \\
\alpha' &= l_{11}y + l_{12}z + l_{13}\alpha + l_{14}\beta, \\
\beta' &= l_{21}y + l_{22}z + l_{23}\alpha + l_{24}\beta,
\end{aligned} \tag{1}
$$

differentiation being with respect to the variable x.

A system of simultaneous solutions $y_i, z_i, \alpha_i, \beta_i (i=1, \cdots, 4)$, with non-vanishing determinant is interpreted geometrically as determining four points P_y, (y_1, y_2, y_3, y_4); P_z, (z_1, z_2, z_3, z_4); P_α, $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$; P_β, $(\beta_1, \beta_2, \beta_3, \beta_4)$, which in turn determine two non-intersecting lines l_{yz}, $l_{\alpha\beta}$. As x varies these points trace four curves C_y, C_z, C_α, C_β, while the lines l_{yz}, $l_{\alpha\beta}$ generate two ruled surfaces R_{yz}, $R_{\alpha\beta}$, on the first of which lie the curves C_y, C_z, and on the second, the curves C_α, C_β.

In this paper we propose to develop the projective differential theory of triads of ruled surfaces whose generators are in one-to-one correspondence. We will determine the defining system of differential equations, calculate certain of the invariants and covariants, and exhibit their geometric significance in a number of instances.

I. The defining system of equations

Let l_{yz}, $l_{\alpha\beta}$ and $l_{\gamma\zeta}$ be three non-intersecting straight lines determined by the respective pairs of points P_y, P_z; P_α, P_β; P_γ, P_ζ; whose coördinates y_i, z_i; α_i, β_i; $\gamma_i, \zeta_i (i=1, \cdots, 4)$, are functions of the independent variable x. It follows that

---

* Presented to the Society, April 3, 1926; received by the editors in April, 1926.

† E. P. Lane, *Ruled surfaces with generators in one-to-one correspondence*, these Transactions, vol. 25 (1923). Hereafter referred to as *Ruled surfaces in correspondence*.

$$(A) \qquad |y_1z_2\alpha_3\beta_4| \neq 0, \quad |\alpha_1\beta_2\gamma_3\zeta_4| \neq 0, \quad |\gamma_1\zeta_2y_3z_4| \neq 0,$$

the determinant in each case being represented by its principal diagonal. As x varies these points trace curves C_y, C_z; C_α, C_β; C_γ, C_ζ; while the lines l_{yz}, $l_{\alpha\beta}$, $l_{\gamma\zeta}$ generate ruled surfaces R_{yz}, $R_{\alpha\beta}$, $R_{\gamma\zeta}$, on which these curves are directrix curves, pair at a time. The generators of these three ruled surfaces will correspond in triples, one set of three for each value of x.

In addition to equations (1) defining the pair of surfaces R_{yz}, $R_{\alpha\beta}$, there will be the two systems,

$$(2) \qquad \begin{aligned} \alpha' &= a_{33}\alpha + a_{34}\beta + a_{35}\gamma + a_{36}\zeta, \\ \beta' &= a_{43}\alpha + a_{44}\beta + a_{45}\gamma + a_{46}\zeta, \\ \gamma' &= m_{33}\alpha + m_{34}\beta + m_{35}\gamma + m_{36}\zeta, \\ \zeta' &= m_{43}\alpha + m_{44}\beta + m_{45}\gamma + m_{46}\zeta, \end{aligned}$$

and

$$(3) \qquad \begin{aligned} \gamma' &= a_{51}y + a_{52}z + a_{55}\gamma + a_{56}\zeta, \\ \zeta' &= a_{61}y + a_{62}z + a_{65}\gamma + a_{66}\zeta, \\ y' &= n_{51}y + n_{52}z + n_{55}\gamma + n_{56}\zeta, \\ z' &= n_{61}y + n_{62}z + n_{65}\gamma + n_{66}\zeta, \end{aligned}$$

defining the pairs $R_{\alpha\beta}$, $R_{\gamma\zeta}$ and $R_{\gamma\zeta}$, R_{yz}.

The system of equations defining the configuration consisting of all three ruled surfaces may be taken to be the combined systems (1), (2), (3). But this system would be cumbersome and involve duplications. As an illustration consider equations (1_1) and (3_3). By elimination of y' we find

$$(4) \qquad \bar{y} = (a_{11} - n_{51})y + (a_{12} - n_{52})z = -(a_{13}\alpha + a_{14}\beta) + (n_{55}\gamma + n_{56}\zeta).$$

Now the point $P_{y'}$, (y_1', y_2', y_3', y_4'), is on the tangent to C_y at P_y and the point $a_{11}y_k + a_{12}z_k (k = 1, \cdots, 4)$ is a point of the line l_{yz}, while the point $a_{13}\alpha_k + a_{14}\beta_k (k = 1, \cdots, 4)$ is on the line $l_{\alpha\beta}$. From these considerations it follows that equation (1_1) expresses analytically the fact that the tangent plane to R_{yz} at P_y cuts the line $l_{\alpha\beta}$ in the point $a_{13}\alpha + a_{14}\beta$*. Similarly equation (3_3) expresses analytically the fact that the same tangent plane cuts $l_{\gamma\zeta}$ in the point $n_{55}\gamma + n_{56}\zeta$. Now the line joining the points $a_{13}\alpha + a_{14}\beta$ and $n_{55}\gamma + n_{56}\zeta$, lying in a plane on l_{yz}, must cut l_{yz}, and equation (4) is the analytic expression of this fact, the point of intersection with l_{yz} being given by the expression $\bar{y} = (a_{11} - n_{51})y + (a_{12} - n_{52})z$. In other words we have expressions for three points on a line of one regulus R_1 of the quadric determined by l_{yz}, $l_{\alpha\beta}$, $l_{\gamma\zeta}$.

---

*By the point $a_{13}\alpha + a_{14}\beta$ we mean the point whose coördinates are $a_{13}\alpha_k + a_{14}\beta_k (k = 1, \cdots, 4)$. This form of abbreviated notation will be employed hereafter.

By making use of equations (1_2) and (3_4) we obtain a second equation

$$\bar{z} = (a_{21} - n_{61})y + (a_{22} - n_{62})z = -(a_{23}\alpha + a_{24}\beta) + (n_{65}\gamma + n_{66}\zeta), \tag{5}$$

giving a second set of three points on a new line of R_1*.

By a linear combination of $\bar{y}$ and $\bar{z}$, $\phi = \kappa\bar{y} + \lambda\bar{z}$, we can obtain any point whatever on l_{yz} and the line of R_1 through this point will cut $l_{\alpha\beta}$ and $l_{\gamma\zeta}$ in the respective points

$$\kappa(a_{13}\alpha + a_{14}\beta) + \lambda(a_{23}\alpha + a_{24}\beta) \text{ and } \kappa(n_{55}\gamma + n_{56}\zeta) + \lambda(n_{65}\gamma + n_{66}\zeta).$$

By a proper choice of κ, λ, we obtain from (4) and (5) a pair of equations of the form

$$\begin{aligned} y &= b_{13}\alpha + b_{14}\beta + b_{15}\gamma + b_{16}\zeta, \\ z &= b_{23}\alpha + b_{24}\beta + b_{25}\gamma + b_{26}\zeta, \end{aligned} \tag{6}$$

the coefficients being of course functions of the coefficients of equations (1_1), (1_2), (3_3), (3_4). Geometrically, equations (6) signify that the line of R_1 through P_y cuts $l_{\alpha\beta}$ and $l_{\gamma\zeta}$ in the respective points $b_{13}\alpha + b_{14}\beta$, $b_{15}\gamma + b_{16}\zeta$, and that the line of R_1 through P_z cuts the lines $l_{\alpha\beta}$, $l_{\gamma\zeta}$ in the respective points $b_{23}\alpha + b_{24}\beta$, $b_{25}\gamma + b_{26}\zeta$. The analytic significance of this situation is that we may replace either pair of equations (1_1), (1_2) or (3_3), (3_4) with the pair (6).

By making use of equations (1_3), (1_4) and (2_1), (2_2) we can obtain a second pair of equations expressing y and z linearly in terms of α, β, γ, ζ. Since there is but one line of R_1 through P_y and but one through P_z, this second pair of equations must be identical with the pair (6). Finally the two pairs (2_3), (2_4) and (3_1), (3_2) again give rise to the pair (6). It follows that if we annex to equations (1), (2), (3), the pair (6), we may eliminate six of the twelve differential equations, thus simplifying greatly the system defining the triad of ruled surfaces $R_{yz}, R_{\alpha\beta}, R_{\gamma\zeta}$. We shall make use of this system in the form

$$(T)\dagger \qquad \begin{aligned} y' &= a_{11}y + a_{12}z + a_{13}\alpha + a_{14}\beta, \\ z' &= a_{21}y + a_{22}z + a_{23}\alpha + a_{24}\beta, \\ \alpha' &= a_{33}\alpha + a_{34}\beta + a_{35}\gamma + a_{36}\zeta, \\ \beta' &= a_{43}\alpha + a_{44}\beta + a_{45}\gamma + a_{46}\zeta, \\ \gamma' &= a_{51}y + a_{52}z + a_{55}\gamma + a_{56}\zeta, \\ \zeta' &= a_{61}y + a_{62}z + a_{65}\gamma + a_{66}\zeta, \\ y &= b_{13}\alpha + b_{14}\beta + b_{15}\gamma + b_{16}\zeta, \\ z &= b_{23}\alpha + b_{24}\beta + b_{25}\gamma + b_{26}\zeta. \end{aligned}$$

*These two lines are in general distinct, since otherwise the tangent planes to R_{yz} at P_y and P_z would coincide, that is, R_{yz} would be a developable surface.

†The first six of these equations constitute what may be termed a semi-canonical form of the general first-order linear system in six dependent variables.

The three sets of equations (1), (2), (3) may be recovered from this simplified system providing only it is possible to solve equations (6) for α, β, and for γ, ζ, that is, if

$$B_1 = b_{13}b_{24} - b_{14}b_{23} \neq 0 \quad \text{and} \quad B_2 = b_{15}b_{26} - b_{16}b_{25} \neq 0.$$

Now $b_{13}b_{24} - b_{14}b_{23} \neq 0$, for otherwise $b_{13} = rb_{23}$, $b_{14} = rb_{24}$, and hence, from (6),

$$y - rz = (b_{15} - rb_{25})\gamma + (b_{16} - rb_{26})\zeta,$$

so that a point $y - rz$ of l_{yz} would be a point $(b_{15} - rb_{25})\gamma + (b_{16} - rb_{26})\zeta$ of $l_{\gamma\zeta}$. But this is contrary to our original assumption. Similarly $b_{15}b_{26} - b_{16}b_{25} \neq 0$.

We note also at this point that the determinants

$$A_1 = a_{13}a_{24} - a_{14}a_{23}, \qquad A_2 = a_{35}a_{46} - a_{36}a_{45}, \qquad A_3 = a_{51}a_{62} - a_{52}a_{61}$$

are non-vanishing. For if, say, $a_{13}a_{24} - a_{14}a_{23} = 0$, then from the first two of equations (T) we would find

$$a_{23}y' - a_{13}z' = (a_{11}a_{23} - a_{21}a_{13})y + (a_{12}a_{23} - a_{22}a_{13})z.$$

But this relation implies that the four points $P_{y'}$, $P_{z'}$, P_y, P_z are coplanar, and hence that R_{yz} is a developable, contrary to our hypothesis. Similarly for the other two determinants.

It is hardly necessary to remark here that for any system of type (T) there always exist twenty-four functions y_i, z_i, α_i, β_i, γ_i, $\zeta_i (i = 1, \cdots, 4)$, linearly independent and satisfying conditions (A) and conversely any such set of twenty-four functions determines a system of equations of type (T).*

The most general transformation which not only preserves the form of system (T) but at the same time leaves our triad of ruled surfaces undisturbed, is

$$\begin{array}{lll} & \bar{x} = \xi(x), & \\ y = c\bar{y} + d\bar{z}, & \alpha = g\bar{\alpha} + h\bar{\beta}, & \gamma = s\bar{\gamma} + t\bar{\zeta}, \\ z = e\bar{y} + f\bar{z}, & \beta = j\bar{\alpha} + k\bar{\beta}, & \zeta = u\bar{\gamma} + v\bar{\zeta}, \\ D_1 = cf - de \neq 0, & D_2 = gk - hj \neq 0, & D_3 = sv - tu \neq 0, \end{array} \tag{7}$$

where the coefficients are functions of x. Geometrically this amounts to the choice of new directrix curves on the three surfaces R_{yz}, $R_{\alpha\beta}$, $R_{\gamma\zeta}$, together with a new parametric representation. If the dependent variables alone are transformed, there results a new system of equations whose coefficients are given by the equations

*All functions are assumed to be analytic.

$$
\begin{aligned}
D_1\bar{a}_{11} &= f(-c' + a_{11}c + a_{12}e) - d(-e' + a_{21}c + a_{22}e),\\
D_1\bar{a}_{12} &= f(-d' + a_{11}d + a_{12}f) - d(-f' + a_{21}d + a_{22}f),\\
D_1\bar{a}_{21} &= -e(-c' + a_{11}c + a_{12}e) + c(-e' + a_{21}c + a_{22}e),\\
D_1\bar{a}_{22} &= -e(-d' + a_{11}d + a_{12}f) + c(-f' + a_{21}d + a_{22}f);
\end{aligned}
\tag{8_1}
$$

$$
\begin{aligned}
D_1\bar{a}_{13} &= f(a_{13}g + a_{14}j) - d(a_{23}g + a_{24}j),\\
D_1\bar{a}_{14} &= f(a_{13}h + a_{14}k) - d(a_{23}h + a_{24}k),\\
D_1\bar{a}_{23} &= -e(a_{13}g + a_{14}j) + c(a_{23}g + a_{24}j),\\
D_1\bar{a}_{24} &= -e(a_{13}h + a_{14}k) + c(a_{23}h + a_{24}k);
\end{aligned}
\tag{8_2}
$$

$$
\begin{aligned}
D_2\bar{a}_{33} &= k(-g' + a_{33}g + a_{34}j) - h(-j' + a_{43}g + a_{44}j),\\
D_2\bar{a}_{34} &= k(-h' + a_{33}h + a_{34}k) - h(-k' + a_{43}h + a_{44}k),\\
D_2\bar{a}_{43} &= -j(-g' + a_{33}g + a_{34}j) + g(-j' + a_{43}g + a_{44}j),\\
D_2\bar{a}_{44} &= -j(-h' + a_{33}h + a_{34}k) + g(-k' + a_{43}h + a_{44}k);
\end{aligned}
\tag{8_3}
$$

$$
\begin{aligned}
D_2\bar{a}_{35} &= k(a_{35}s + a_{36}u) - h(a_{45}s + a_{46}u),\\
D_2\bar{a}_{36} &= k(a_{35}t + a_{36}v) - h(a_{45}t + a_{46}v),\\
D_2\bar{a}_{45} &= -j(a_{35}s + a_{36}u) + g(a_{45}s + a_{46}u),\\
D_2\bar{a}_{46} &= -j(a_{35}t + a_{36}v) + g(a_{45}t + a_{46}v);
\end{aligned}
\tag{8_4}
$$

$$
\begin{aligned}
D_3\bar{a}_{55} &= v(-s' + a_{55}s + a_{56}u) - t(-u' + a_{65}s + a_{66}u),\\
D_3\bar{a}_{56} &= v(-t' + a_{55}t + a_{56}v) - t(-v' + a_{65}t + a_{66}v),\\
D_3\bar{a}_{65} &= -u(-s' + a_{55}s + a_{56}u) + s(-u' + a_{65}s + a_{66}u),\\
D_3\bar{a}_{66} &= -u(-t' + a_{55}t + a_{56}v) + s(-v' + a_{65}t + a_{66}v);
\end{aligned}
\tag{8_5}
$$

$$
\begin{aligned}
D_3\bar{a}_{51} &= v(a_{51}c + a_{52}e) - t(a_{61}c + a_{62}e),\\
D_3\bar{a}_{52} &= v(a_{51}d + a_{52}f) - t(a_{61}d + a_{62}f),\\
D_3\bar{a}_{61} &= -u(a_{51}c + a_{52}e) + s(a_{61}c + a_{62}e),\\
D_3\bar{a}_{62} &= -u(a_{51}d + a_{52}f) + s(a_{61}d + a_{62}f);
\end{aligned}
\tag{8_6}
$$

$$
\begin{aligned}
D_3\bar{b}_{13} &= f(b_{13}g + b_{14}j) - d(b_{23}g + b_{24}j),\\
D_3\bar{b}_{14} &= f(b_{13}h + b_{14}k) - d(b_{23}h + b_{24}k),\\
D_3\bar{b}_{23} &= -e(b_{13}g + b_{14}j) + c(b_{23}g + b_{24}j),\\
D_3\bar{b}_{24} &= -e(b_{13}h + b_{14}k) + c(b_{23}h + b_{24}k);
\end{aligned}
\tag{8_7}
$$

$$
\begin{aligned}
D_1\bar{b}_{15} &= f(b_{15}s + b_{16}u) - d(b_{25}s + b_{26}u),\\
D_1\bar{b}_{16} &= f(b_{15}t + b_{16}v) - d(b_{25}t + b_{26}v),\\
D_1\bar{b}_{25} &= -e(b_{15}s + b_{16}u) + c(b_{25}s + b_{26}u),\\
D_1\bar{b}_{26} &= -e(b_{15}t + b_{16}v) + c(b_{25}t + b_{26}v).
\end{aligned}
\tag{8_8}
$$

From (8_1) it follows that if c, d, e, f are chosen as a set of simultaneous solutions of the system of differential equations

$$c' = a_{11}c + a_{12}e,\quad d' = a_{11}d + a_{12}f,\quad e' = a_{21}c + a_{22}e,\quad f' = a_{21}d + a_{22}f,$$

then for our new system we shall have $\bar{a}_{11} = \bar{a}_{12} = \bar{a}_{21} = \bar{a}_{22} = 0$. The new curves $C_{\bar{y}}$, $C_{\bar{z}}$, on R_{yz}, thus obtained, have been called intersector curves by Lane*, from the fact that their tangents at $P_{\bar{y}}$, $P_{\bar{z}}$, on l_{yz} intersect the corresponding line $l_{\alpha\beta}$ on $R_{\alpha\beta}$.

In a similar manner we may choose g, h, j, k as a set of solutions of the system

$$g' = a_{33}g + a_{34}j,\quad h' = a_{33}h + a_{34}k,\quad j' = a_{43}g + a_{44}j,\quad k' = a_{43}h + a_{44}k,$$

and s, t, u, v, a set of solutions of the system

$$s' = a_{55}s + a_{56}u,\quad t' = a_{55}t + a_{56}v,\quad u' = a_{65}s + a_{66}u,\quad v' = a_{65}t + a_{66}v,$$

and thus reduce our system (T) to the canonical form

$$
\begin{aligned}
y' &= a_{13}\alpha + a_{14}\beta,\quad \alpha' = a_{35}\gamma + a_{36}\zeta,\quad \gamma' = a_{51}y + a_{52}z,\\
z' &= a_{23}\alpha + a_{24}\beta,\quad \beta' = a_{45}\gamma + a_{46}\zeta,\quad \zeta' = a_{61}y + a_{62}z,\\
y &= b_{13}\alpha + b_{14}\beta + b_{15}\gamma + b_{16}\zeta,\\
z &= b_{23}\alpha + b_{24}\beta + b_{25}\gamma + b_{26}\zeta,
\end{aligned}
\tag{R}
$$

where for simplicity we write a_{ij}, $b_{\kappa\lambda}$, instead of $\bar{a}_{ij}$, $\bar{b}_{\kappa\lambda}$.

For this system the directrix curves on surface R_{yz} are intersector curves with respect to the surface $R_{\alpha\beta}$, the directrix curves on $R_{\alpha\beta}$ are intersector curves with respect to $R_{\gamma\zeta}$, and so on. This is apparent from the first six of the above equations. Lane† has called attention to the fact that for two ruled surfaces whose elements are in one-to-one correspondence, there exists on each a one-parameter family of curves which are intersector curves with respect to the other. This follows when we note that the form of system (R) remains unchanged so long as the dependent variables are subjected to trans-

---

* *Ruled surfaces in correspondence.*

† *Ruled surfaces in correspondence.*

formations involving constant coefficients. It must be noted, however, that the directrix curves on R_{yz} are not intersector curves with respect to the surface $R_{\gamma\zeta}$, nor the directrix curves on $R_{\gamma\zeta}$ intersector curves with respect to $R_{\alpha\beta}$. If we wish to reverse the order we must first rewrite system (T) in such a way that y', z' are expressed in terms of $y, z, \gamma, \zeta; \gamma', \zeta'$ in terms of $\gamma, \zeta, \alpha, \beta$, and α', β', in terms of α, β, y, z. This change may be effected easily by the use of the last two equations of the system. When this has been done the new system may then be reduced to its canonical form by suitably chosen transformations on the dependent variables. It is unnecessary to supply the details at this time.

In the sequel we shall have occasion to make use of cyclic substitutions on the dependent variables and the coefficients of systems (T) and (R) corresponding to the three-fold symmetry of our configuration. For either system, if the order of the first six equations is

$$\alpha' = \cdots, \quad \gamma' = \cdots, \quad y' = \cdots,$$
$$\beta' = \cdots, \quad \zeta' = \cdots, \quad z' = \cdots,$$

then logically the last pair should be solved for α, β, thus,

$$\begin{aligned} \alpha &= [(b_{14}b_{25} - b_{15}b_{24})\gamma + (b_{14}b_{26} - b_{16}b_{24})\zeta + b_{24}y - b_{14}z]/B_1, \\ \beta &= [-(b_{13}b_{25} - b_{15}b_{23})\gamma - (b_{13}b_{26} - b_{16}b_{23})\zeta - b_{23}y + b_{13}z]/B_1, \end{aligned} \tag{9}$$

and if the order of the first six is

$$\gamma' = \cdots, \quad y' = \cdots, \quad \alpha' = \cdots,$$
$$\zeta' = \cdots, \quad z' = \cdots, \quad \beta' = \cdots,$$

the last pair should be solved for γ, ζ,

$$\begin{aligned} \gamma &= [b_{26}y - b_{16}z - (b_{13}b_{26} - b_{16}b_{23})\alpha - (b_{14}b_{26} - b_{16}b_{24})\beta]/B_2, \\ \zeta &= [-b_{25}y + b_{15}z + (b_{13}b_{25} - b_{15}b_{23})\alpha + (b_{14}b_{25} - b_{15}b_{24})\beta]/B_2. \end{aligned} \tag{10}$$

The corresponding cyclic substitution on the dependent variables is $(y\alpha\gamma)$ $(z\beta\zeta)$. Permuting the a_{ij} replaces every a_{ij} with $a_{i+2,j+2}$, it being understood that wherever $i+2$ or $j+2$ exceeds 6, we use its residue, mod 6. The permutations on the b's involve the replacing of the coefficients in the last two equations of (T) or (R) by those in (9) taken in the same order, and those in (9) by those in (10).

Before proceeding to the calculation of some of the invariants and covariants of our system of equations, it is worth while to point out that the derivatives of the coefficients $b_{\kappa\lambda}$ are functions of a_{ij}, $b_{\kappa\lambda}$. If the last pair of equations in either (T) or (R) be differentiated we shall find thus two ex-

pressions for y' and two for z'. By equating these values and making use of the remaining equations, there arise two identities in $y, z, \alpha, \beta, \gamma, \zeta$. Equating coefficients enables us to write down the values of $b'_{\kappa\lambda}$ in terms of a_{ij}, $b_{\kappa\lambda}$.

II. Invariants and covariants

The infinitesimal transformations of the dependent variables may be written

$$
\begin{aligned}
&y = (1+\phi_1\delta t)\bar{y} + \psi_1\delta t\bar{z}, \qquad z = \kappa_1\delta t\bar{y} + (1+\omega_1\delta t)\bar{z}\ ;\\
&\alpha = (1+\phi_2\delta t)\bar{\alpha} + \psi_2\delta t\bar{\beta}, \qquad \beta = \kappa_2\delta t\bar{\alpha} + (1+\omega_2\delta t)\bar{\beta}\ ;\\
&\gamma = (1+\phi_3\delta t)\bar{\gamma} + \psi_3\delta t\bar{\zeta}, \qquad \zeta = \kappa_3\delta t\bar{\gamma} + (1+\omega_3\delta t)\bar{\zeta}\ ;
\end{aligned} \tag{11}
$$

$$
\begin{aligned}
D_1 &= 1 + (\phi_1 + \omega_1)\delta t\ ;\\
D_2 &= 1 + (\phi_2 + \omega_2)\delta t\ ;\\
D_3 &= 1 + (\phi_3 + \omega_3)\delta t\ ;
\end{aligned}
$$

where ϕ_i, ψ_i, κ_i, $\omega_i (i=1, 2, 3)$ are arbitrary functions of x, and t is independent of x.

From (8) and (11) we find for the infinitesimal changes in the coefficients of system (T),

$$
\begin{aligned}
\delta a_{11} &= (-\phi_1' + a_{12}\kappa_1 - a_{21}\psi_1)\delta t,\\
\delta a_{12} &= [-\psi_1' + (a_{11} - a_{22})\psi_1 - a_{12}(\phi_1 - \omega_1)]\delta t,\\
\delta a_{21} &= [-\kappa_1' - (a_{11} - a_{22})\kappa_1 + a_{21}(\phi_1 - \omega_1)]\delta t,\\
\delta a_{22} &= (-\omega_1' - a_{12}\kappa_1 + a_{21}\psi_1)\delta t\ ;
\end{aligned} \tag{12$_1$}
$$

$$
\begin{aligned}
\delta a_{13} &= [a_{13}(\phi_2 - \phi_1) + a_{14}\kappa_2 - a_{23}\psi_1]\delta t,\\
\delta a_{14} &= [a_{13}\psi_2 + a_{14}(\omega_2 - \phi_1) - a_{24}\psi_1]\delta t,\\
\delta a_{23} &= [-a_{13}\kappa_1 + a_{23}(\phi_2 - \omega_1) + a_{24}\kappa_2]\delta t,\\
\delta a_{24} &= [-a_{14}\kappa_1 + a_{23}\psi_2 + a_{24}(\omega_2 - \omega_1)]\delta t\ ;
\end{aligned} \tag{12$_2$}
$$

$$
\begin{aligned}
\delta a_{33} &= (-\phi_2' + a_{34}\kappa_2 - a_{43}\psi_2)\delta t,\\
\delta a_{34} &= [-\psi_2' + (a_{33} - a_{44})\psi_2 - a_{34}(\phi_2 - \omega_2)]\delta t,\\
\delta a_{43} &= [-\kappa_2' - (a_{33} - a_{44})\kappa_2 + a_{43}(\phi_2 - \omega_2)]\delta t,\\
\delta a_{44} &= (-\omega_2' - a_{34}\kappa_2 + a_{43}\psi_2)\delta t\ ;
\end{aligned} \tag{12$_3$}
$$

$$
\begin{aligned}
\delta a_{35} &= [a_{35}(\phi_3 - \phi_2) + a_{36}\kappa_3 - a_{45}\psi_2]\delta t,\\
\delta a_{36} &= [a_{35}\psi_3 + a_{36}(\omega_3 - \phi_2) - a_{46}\psi_2]\delta t,\\
\delta a_{45} &= [-a_{35}\kappa_2 + a_{45}(\phi_3 - \omega_2) + a_{46}\kappa_3]\delta t,\\
\delta a_{46} &= [-a_{36}\kappa_2 + a_{45}\psi_3 + a_{46}(\omega_3 - \omega_2)]\delta t\ ;
\end{aligned} \tag{12$_4$}
$$

$$
(12_5)\qquad
\begin{aligned}
\delta a_{55} &= (-\phi_3' + a_{56}\kappa_3 - a_{65}\psi_3)\delta t,\\
\delta a_{56} &= [-\psi_3' + (a_{55} - a_{66})\psi_3 - a_{56}(\phi_3 - \omega_3)]\delta t,\\
\delta a_{65} &= [-\kappa_3' - (a_{55} - a_{66})\kappa_3 + a_{65}(\phi_3 - \omega_3)]\delta t,\\
\delta a_{66} &= (-\omega_3' - a_{56}\kappa_3 + a_{65}\psi_3)\delta t\ ;
\end{aligned}
$$

$$
(12_6)\qquad
\begin{aligned}
\delta a_{51} &= [a_{51}(\phi_1 - \phi_3) + a_{52}\kappa_1 - a_{61}\psi_3]\delta t,\\
\delta a_{52} &= [a_{51}\psi_1 + a_{52}(\omega_1 - \phi_3) - a_{62}\psi_3]\delta t,\\
\delta a_{61} &= [-a_{51}\kappa_3 + a_{61}(\phi_1 - \omega_3) + a_{62}\kappa_1]\delta t,\\
\delta a_{62} &= [-a_{52}\kappa_3 + a_{61}\psi_1 + a_{62}(\omega_1 - \omega_3)]\delta t\ ;
\end{aligned}
$$

$$
(12_7)\qquad
\begin{aligned}
\delta b_{13} &= [b_{13}(\phi_2 - \phi_1) + b_{14}\kappa_2 - b_{23}\psi_1]\delta t,\\
\delta b_{14} &= [b_{13}\psi_2 + b_{14}(\omega_2 - \phi_1) - b_{24}\psi_1]\delta t,\\
\delta b_{23} &= [-b_{13}\kappa_1 + b_{23}(\phi_2 - \omega_1) + b_{24}\kappa_2]\delta t,\\
\delta b_{24} &= [-b_{14}\kappa_1 + b_{23}\psi_2 + b_{24}(\omega_2 - \omega_1)]\delta t\ ;
\end{aligned}
$$

$$
(12_8)\qquad
\begin{aligned}
\delta b_{15} &= [b_{15}(\phi_3 - \phi_1) + b_{16}\kappa_3 - b_{25}\psi_1]\delta t,\\
\delta b_{16} &= [b_{15}\psi_3 + b_{16}(\omega_3 - \phi_1) - b_{26}\psi_1]\delta t,\\
\delta b_{25} &= [-b_{15}\kappa_1 + b_{25}(\phi_3 - \omega_1) + b_{26}\kappa_3]\delta t,\\
\delta b_{26} &= [-b_{16}\kappa_1 + b_{25}\psi_3 + b_{26}(\omega_3 - \omega_1)]\delta t.
\end{aligned}
$$

Any function $U(a_{ij}, b_{\kappa\lambda})$, invariant under transformations on the dependent variables, must be such that

$$\delta U = \sum \frac{\partial U}{\partial a_{ij}}\delta a_{ij} + \sum \frac{\partial U}{\partial b_{\kappa\lambda}}\delta b_{\kappa\lambda} = 0.$$

If we introduce into this equation the values obtained above and equate to zero the coefficients of $\phi_i, \phi_i', \psi_i, \psi_i', \kappa_i, \kappa_i', \omega_i, \omega_i'$ $(i=1, 2, 3)$, we will obtain a system of partial differential equations whose solutions will be absolute seminvariants of system (T).

Without following the investigation further at this time we wish to remark here that the coefficients of $\phi_1', \psi_1', \kappa_1', \omega_1'$, when equated to zero, give

$$\frac{\partial U}{\partial a_{11}} = 0,\quad \frac{\partial U}{\partial a_{12}} = 0,\quad \frac{\partial U}{\partial a_{21}} = 0,\quad \frac{\partial U}{\partial a_{22}} = 0,$$

so that *there are no seminvariants containing* a_{11}, a_{12}, a_{21}, a_{22}. Similarly it follows that *there are no seminvariants containing* a_{33}, a_{34}, a_{43}, a_{44}; a_{55}, a_{56}, a_{65}, a_{66}. Since the only coefficients which can appear in the absolute seminvariants are tbose of the canonical form (R), we conclude that *the canonical and*

uncanonical forms of the absolute seminvariants are identical. In our search for invariants we may therefore confine our attention to system (R).

Referring to equations (8_2), (8_4), (8_6), we discover that they constitute an invariant system under cyclic substitution on the coefficients a_{ij} providing also we insist on subjecting the coefficients of (7) to the cyclic substitution $(cgs)(dht)(eju)(fkv)$. Further investigation discloses that the two sets (8_7) and (8_8) possess the same property. In view of this fact we find that *if any function $U(a_{ij}, b_{\kappa\lambda})$ is a seminvariant, then the new function obtained from it by permuting the arguments of a_{ij}, $b_{\kappa\lambda}$ is also a seminvariant.*

We are ready now to calculate some of the seminvariants. If the four equations (12_6) be multiplied respectively by b_{15}, b_{25}, b_{16}, b_{26}, and the four equations (12_8) similarly by a_{51}, a_{61}, a_{52}, a_{62}, and the resulting equations added, we find that

$$a_{51}\delta b_{15} + b_{15}\delta a_{51} + a_{52}\delta b_{25} + b_{25}\delta a_{52} + a_{61}\delta b_{16} + b_{16}\delta a_{61} + a_{62}\delta b_{26} + b_{26}\delta a_{62} = 0.$$

We have thus found an absolute seminvariant

$$I_1 = a_{51}b_{15} + a_{52}b_{25} + a_{61}b_{16} + a_{62}b_{26}. \tag{13}$$

From this, by permuting the letters we obtain two more seminvariants. They are

$$I_1^{(1)} = (a_{13}b_{24} - a_{14}b_{23} - a_{23}b_{14} + a_{24}b_{13})/{}_1B, \tag{14}$$

and

$$I_1^{(2)} = [- a_{35}(b_{13}b_{26} - b_{16}b_{23}) + a_{36}(b_{13}b_{25} - b_{15}b_{23}) - a_{45}(b_{14}b_{26} - b_{16}b_{24}) + a_{46}(b_{14}b_{25} - b_{15}b_{24})]/B_2. \tag{15}$$

For the five determinants A_1, A_2, A_3, B_1, B_2 we find from equations (12)

$$\begin{aligned} \delta A_1 &= A_1(- \phi_1 + \phi_2 - \omega_1 + \omega_2)\delta t, \\ \delta A_2 &= A_2(- \phi_2 + \phi_3 - \omega_2 + \omega_3)\delta t, \\ \delta A_3 &= A_3(- \phi_3 + \phi_1 - \omega_3 + \omega_1)\delta t, \\ \delta B_1 &= B_1(- \phi_1 + \phi_2 - \omega_1 + \omega_2)\delta t, \\ \delta B_2 &= B_2(- \phi_1 + \phi_3 - \omega_1 + \omega_3)\delta t. \end{aligned} \tag{16}$$

From (16) we find

$$A_3\delta B_2 + B_2\delta A_3 = 0,$$

so that

$$I_2 = A_3B_2 = (a_{51}a_{62} - a_{52}a_{61})(b_{15}b_{26} - b_{16}b_{25}) \tag{17}$$

is an absolute seminvariant. Again from (16) we have

$$B_1\delta A_1 - A_1\delta B_1 = 0,$$

$$B_2(A_2\delta B_1 + B_1\delta A_2) - A_2B_1\delta B_2 = 0,$$

so that

$$I_2^{(1)} = A_1/B_1 = (a_{13}a_{24} - a_{14}a_{23})/(b_{13}b_{24} - b_{14}b_{23}) \tag{18}$$

and

$$I_2^{(2)} = A_2B_1/B_2 = (a_{35}a_{46} - a_{36}a_{45})(b_{13}b_{24} - b_{14}b_{23})/(b_{15}b_{26} - b_{16}b_{25}) \tag{19}$$

are also absolute seminvariants. $I_2^{(1)}$ and $I_2^{(2)}$ might have been obtained from I_2 by permuting the letters.

It may be verified by reference to equations (8) or (12) that

$$J_2 = b_{13}(a_{35}a_{51} + a_{36}a_{61}) + b_{14}(a_{45}a_{51} + a_{46}a_{61}) + b_{23}(a_{35}a_{52} + a_{36}a_{62}) + b_{24}(a_{45}a_{52} + a_{46}a_{62}) \tag{20}$$

is an absolute seminvariant. From this, by permuting letters, we find two new seminvariants. They are $J_2^{(1)}$ and $J_2^{(2)}$ where

$$B_1J_2^{(1)} = (a_{13}a_{51} + a_{23}a_{52})(b_{14}b_{25} - b_{15}b_{24}) + (a_{13}a_{61} + a_{23}a_{62})(b_{14}b_{26} - b_{16}b_{24}) - (a_{14}a_{51} + a_{24}a_{52})(b_{13}b_{25} - b_{15}b_{23}) - (a_{14}a_{61} + a_{24}a_{62})(b_{13}b_{26} - b_{16}b_{23}), \tag{21}$$

$$B_2J_2^{(2)} = b_{15}(a_{23}a_{36} + a_{24}a_{46}) - b_{16}(a_{23}a_{35} + a_{24}a_{45}) - b_{25}(a_{13}a_{36} + a_{14}a_{46}) + b_{26}(a_{13}a_{35} + a_{14}a_{45}). \tag{22}$$

Let $Kb_{13}+Lb_{14}+Mb_{23}+Nb_{24}$ and $Pb_{13}+Qb_{14}+Rb_{23}+Sb_{24}$ be any two relative seminvariants linearly expressible in terms of b_{13}, b_{14}, b_{23}, b_{24}, and whose coefficients are functions of a_{ij}. Then we have

$$\mu(\bar{K}\bar{b}_{13} + \bar{L}\bar{b}_{14} + \bar{M}\bar{b}_{23} + \bar{N}\bar{b}_{24}) = Kb_{13} + Lb_{14} + Mb_{23} + Nb_{24},$$

$$\nu(\bar{P}\bar{b}_{13} + \bar{Q}\bar{b}_{14} + \bar{R}\bar{b}_{23} + \bar{S}\bar{b}_{24}) = Pb_{13} + Qb_{14} + Rb_{23} + Sb_{24}.$$

Replacing $\bar{b}_{13}, \cdots, \bar{b}_{24}$ by their values from (8_7) and equating coefficients of $b_{13}, \cdots, b_{24}$, on opposite sides of these identities, we find

$$D_1K = \mu(fg\bar{K} + fh\bar{L} - eg\bar{M} - eh\bar{N}),$$

$$D_1L = \mu(fj\bar{K} + fk\bar{L} - ej\bar{M} - ek\bar{N}),$$

$$D_1M = \mu(-dg\bar{K} - dh\bar{L} + cg\bar{M} + ch\bar{N}),$$

$$D_1N = \mu(-dj\bar{K} - dk\bar{L} + cj\bar{M} + ck\bar{N});$$

$$D_1P = \nu(fg\bar{P} + fh\bar{Q} - eg\bar{R} - eh\bar{S}),$$
$$D_1Q = \nu(fj\bar{P} + fk\bar{Q} - ej\bar{R} - ek\bar{S}),$$
$$D_1R = \nu(-dg\bar{P} - dh\bar{Q} + cg\bar{R} + ch\bar{S}),$$
$$D_1S = \nu(-dj\bar{P} - dk\bar{Q} + cj\bar{R} + ck\bar{S}),$$

and from these relations it follows that

$$D_1(KS - LR - MQ + NP) = \mu\nu D_2(\bar{K}\bar{S} - \bar{L}\bar{R} - \bar{M}\bar{Q} + \bar{N}\bar{P}),$$

that is, $KS-LR-MQ+NP$ is a relative seminvariant.*

Now the numerator $H = a_{13}b_{24} - a_{14}b_{23} - a_{23}b_{14} + a_{24}b_{13}$ of the seminvariant $I_1^{(1)}$ can be shown to be a relative seminvariant for which $D_2\mu = D_1$. Using H and the seminvariant J_2 for which $\nu = 1$, we find by the above theorem that

$$(23)\quad J_3 = a_{13}(a_{35}a_{51} + a_{36}a_{61}) + a_{14}(a_{45}a_{51} + a_{46}a_{61}) + a_{23}(a_{35}a_{52} + a_{36}a_{62}) + a_{24}(a_{45}a_{52} + a_{46}a_{62})$$

is a new absolute seminvariant. Permuting the letters of J_3 gives J_3 itself.

We have now found ten absolute seminvariants, four of which are integral in form. If we wish a set of ten, all of which are integral, we may take $I_1, I_2, J_2, I_3 = I_2I_1^{(2)}, J_3, I_4 = I_2, I_2^{(2)}, J_4 = I_2J_2^{(2)}, I_5 = I_2I_1^{(1)}I_2^{(2)}, I_6 = I_2I_2^{(1)}I_2^{(2)}, J_6 = I_2I_2^{(2)}J_2^{(1)}$. Of these ten, I_2, I_4, I_6 cannot vanish identically.

If, following the usual custom, we assign to a_{ij} the weight one and to $b_{\kappa\lambda}$ the weight zero, then the subscripts above indicate the weights of the seminvariants.

A transformation $\bar{x} = \xi(x)$, of the independent variable, is seen to replace every coefficient a_{ij} with a_{ij}/ξ' and to leave the coefficients $b_{\kappa\lambda}$ unchanged. It follows that *the ten absolute seminvariants are all relative invariants.*

As a preliminary to the search for covariants we wish to establish a theorem which will be of great assistance to us. Let $Py+Qz$ and $Ry+Sz$ be any two relative semicovariants, linear in y and z, so that

$$\mu(\bar{P}\bar{y} + \bar{Q}\bar{z}) = Py + Qz, \qquad \nu(\bar{R}\bar{y} + \bar{S}\bar{z}) = Ry + Sz.$$

We have, by (7),

$$\mu(\bar{P}\bar{y} + \bar{Q}\bar{z}) = \mu[(\bar{P}c + \bar{Q}e)y + (\bar{P}d + \bar{Q}f)z] = Py + Qz,$$
$$\nu(\bar{R}\bar{y} + \bar{S}\bar{z}) = \nu[(\bar{R}c + \bar{S}e)y + (\bar{R}d + \bar{S}f)z] = Ry + Sz,$$

---

*$b_{15}, b_{16}, b_{25}, b_{26}$ might have been used rather than $b_{13}, b_{14}, b_{23}, b_{24}$, without altering the conclusion.

so that

$$\mu(\overline{P}c + \overline{Q}e) = P, \qquad \mu(\overline{P}d + \overline{Q}f) = Q,$$
$$\nu(\overline{R}c + \overline{S}e) = R, \qquad \nu(\overline{R}d + \overline{S}f) = S,$$

and therefore

$$PS - QR = \mu\nu D_1(\overline{P}\overline{S} - \overline{Q}\overline{R}).$$

It follows that *the determinant of the coefficients of y and z in any two relative semicovariants, linear in y and z, is a relative semicovariant, or relative seminvariant, according as these coefficients are, or are not, functions of the dependent variables.* This theorem holds if y and z are replaced by either of the pairs $\alpha\beta$; $\gamma\zeta$.

By making use of equations (7) and (8) we find that

$$(24) \qquad \begin{aligned} C_0^{(1)} &= (b_{23}y - b_{13}z)\alpha + (b_{24}y - b_{14}z)\beta, \\ C_1^{(1)} &= (a_{23}y - a_{13}z)\alpha + (a_{24}y - a_{14}z)\beta \end{aligned}$$

are relative semicovariants, since $D_1\overline{C}_0^{(1)} = C_0^{(1)}$ and $D_1\overline{C}_1^{(1)} = C_1^{(1)}$. By introducing into the first of (24) the values of α and β given in (9), or the values of y and z given in (R), we obtain two alternative forms of $C_0^{(1)}$. They are

$$C_0^{(2)} = (b_{25}y - b_{15}z)\gamma + (b_{26}y - b_{16}z)\zeta \qquad (C_0^{(2)} = -C_0^{(1)}),$$

$$C_0^{(3)} = [(b_{13}b_{25} - b_{15}b_{23})\alpha + (b_{14}b_{25} - b_{15}b_{24})\beta]\gamma + [(b_{13}b_{26} - b_{16}b_{23})\alpha + (b_{14}b_{26} - b_{16}b_{24})\beta]\zeta \qquad (C_0^{(3)} = C_0^{(1)}).$$

By permuting the letters we obtain from the second of equations (24) two additional relative semicovariants. They are

$$(25) \qquad \begin{aligned} C_1^{(2)} &= (a_{45}\alpha - a_{35}\beta)\gamma + (a_{46}\alpha - a_{36}\beta)\zeta, \\ C_1^{(3)} &= (a_{61}\gamma - a_{51}\zeta)y + (a_{62}\gamma - a_{52}\zeta)z, \end{aligned}$$

for which $D_2\overline{C}_1^{(2)} = C_1^{(2)}$ and $D_3\overline{C}_1^{(3)} = C_1^{(3)}$.

Since $C_0^{(1)}$ and $C_1^{(1)}$ are expressible as both linear in y, z, or both linear in α, β, we obtain from them, by the theorem proved above, two new relative semicovariants. They are

$$(26) \qquad \begin{aligned} K_1^{(1)} &= (a_{23}b_{24} - a_{24}b_{23})y^2 - (a_{13}b_{24} - a_{14}b_{23} + a_{23}b_{14} - a_{24}b_{13})yz \\ &\qquad + (a_{13}b_{14} - a_{14}b_{13})z^2 \qquad (D_1^2\,\overline{K}_1^{(1)} = D_2K_1^{(1)}), \\ K_1^{(2)} &= (a_{13}b_{23} - a_{23}b_{13})\alpha^2 + (a_{13}b_{24} + a_{14}b_{23} - a_{23}b_{14} - a_{24}b_{13})\alpha\beta \\ &\qquad + (a_{14}b_{24} - a_{24}b_{14})\beta^2 \qquad (D_1\,\overline{K}_1^{(2)} = K_1^{(2)}). \end{aligned}$$

Similarly, from $C_0^{(2)}$ and $C_1^{(3)}$ we obtain

$$
\begin{aligned}
L_1^{(1)} &= (a_{51}b_{25} + a_{61}b_{26})y^2 - (a_{51}b_{15} - a_{52}b_{25} + a_{61}b_{16} - a_{62}b_{26})yz \\
&\qquad - (a_{52}b_{15} + a_{62}b_{16})z^2 \qquad (D_1\bar{L}_1^{(1)} = L_1^{(1)}), \\
L_1^{(2)} &= (a_{61}b_{15} + a_{62}b_{25})\gamma^2 - (a_{51}b_{15} + a_{52}b_{25} - a_{61}b_{16} - a_{62}b_{26})\gamma\zeta \\
&\qquad - (a_{51}b_{16} + a_{52}b_{26})\zeta^2 \qquad (D_3\bar{L}_1^{(2)} = L_1^{(2)}),
\end{aligned} \tag{27}
$$

and from $C_0^{(3)}$ and $C_1^{(2)}$,

$$
\begin{aligned}
M_1^{(1)} &= [a_{45}(b_{13}b_{26} - b_{16}b_{23}) - a_{46}(b_{13}b_{25} - b_{15}b_{23})]\alpha^2 \\
&\quad - [a_{35}(b_{13}b_{26} - b_{16}b_{23}) - a_{36}(b_{13}b_{25} - b_{15}b_{23}) - a_{45}(b_{14}b_{26} - b_{16}b_{24}) \\
&\quad + a_{46}(b_{14}b_{25} - b_{15}b_{24})]\alpha\beta - [a_{35}(b_{14}b_{26} - b_{16}b_{24}) \\
&\quad - a_{36}(b_{14}b_{25} - b_{15}b_{24})]\beta^2, \\
M_1^{(2)} &= [a_{35}(b_{13}b_{25} - b_{15}b_{23}) + a_{45}(b_{14}b_{25} - b_{15}b_{24})]\gamma^2 \\
&\quad + [a_{35}(b_{13}b_{26} - b_{16}b_{23}) + a_{36}(b_{13}b_{25} - b_{15}b_{23}) \\
&\quad + a_{45}(b_{14}b_{26} - b_{16}b_{24}) + a_{46}(b_{14}b_{25} - b_{15}b_{24})]\gamma\zeta \\
&\quad + [a_{36}(b_{13}b_{26} - b_{16}b_{23}) + a_{46}(b_{14}b_{26} - b_{16}b_{24})]\zeta^2,
\end{aligned} \tag{28}
$$

where

$$D_1D_2\bar{M}_1^{(1)} = D_3M_1^{(1)}, \qquad D_1\bar{M}_1^{(2)} = M_1^{(2)}.$$

Writing $C_1^{(2)}$ in the form $(a_{45}\gamma + a_{46}\zeta)\alpha - (a_{35}\gamma + a_{36}\zeta)\beta$, we obtain from it and $C_1^{(1)}$ by the process employed above

$$
\begin{aligned}
C_2^{(1)} &= [(a_{23}a_{35} + a_{24}a_{45})\gamma + (a_{23}a_{36} + a_{24}a_{46})\zeta]y - [(a_{13}a_{35} + a_{14}a_{45})\gamma \\
&\qquad + (a_{13}a_{36} + a_{14}a_{46})\zeta]z \qquad (D_1\bar{C}_2^{(1)} = C_2^{(1)}).
\end{aligned} \tag{29}
$$

By repeating this process with $C_2^{(1)}$ and $C_1^{(3)}$ we obtain

$$
\begin{aligned}
C_3^{(3)} &= (a_{13}a_{35}a_{61} + a_{14}a_{45}a_{61} + a_{23}a_{35}a_{62} + a_{24}a_{45}a_{62})\gamma^2 \\
&\quad - [a_{13}(a_{35}a_{51} - a_{36}a_{61}) + a_{14}(a_{45}a_{51} - a_{46}a_{61}) + a_{23}(a_{35}a_{52} - a_{36}a_{62}) \\
&\quad + a_{24}(a_{45}a_{52} - a_{46}a_{62})]\gamma\zeta - (a_{13}a_{36}a_{51} + a_{14}a_{46}a_{51} + a_{23}a_{36}a_{52} \\
&\quad + a_{24}a_{46}a_{52})\zeta^2.
\end{aligned} \tag{30}
$$

By rearranging $C_2^{(1)}$ and $C_1^{(3)}$ the determinant of the coefficients of γ, ζ is

$$
\begin{aligned}
C_3^{(1)} &= (a_{35}a_{51}a_{23} + a_{36}a_{61}a_{23} + a_{45}a_{51}a_{24} + a_{46}a_{61}a_{24})y^2 \\
&\quad - [a_{35}(a_{51}a_{13} - a_{52}a_{23}) + a_{36}(a_{61}a_{13} - a_{62}a_{23}) + a_{45}(a_{51}a_{14} - a_{52}a_{24}) \\
&\quad + a_{46}(a_{61}a_{14} - a_{62}a_{24})]yz - (a_{35}a_{52}a_{13} + a_{36}a_{62}a_{13} + a_{45}a_{52}a_{14} \\
&\quad + a_{46}a_{62}a_{14})z^2.
\end{aligned} \tag{31}
$$

By comparison of $C_3^{(3)}$ and $C_3^{(1)}$ it is seen that the latter can be obtained from the former by permuting the letters. This suggests that we permute the letters in $C_3^{(1)}$. When this is done there results

$$
\begin{aligned}
C_3^{(2)} = {} & (a_{51}a_{13}a_{45} + a_{52}a_{23}a_{45} + a_{61}a_{13}a_{46} + a_{62}a_{23}a_{46})\alpha^2 \\
& - [a_{51}(a_{13}a_{35} - a_{14}a_{45}) + a_{52}(a_{23}a_{35} - a_{24}a_{45}) + a_{61}(a_{13}a_{36} - a_{14}a_{46}) \\
& + a_{62}(a_{23}a_{36} - a_{24}a_{46})]\alpha\beta - (a_{51}a_{14}a_{35} + a_{52}a_{24}a_{35} \\
& + a_{61}a_{14}a_{36} + a_{62}a_{24}a_{36})\beta^2.
\end{aligned} \tag{32}
$$

For these last three relative semicovariants we find

$$D_1\overline{C}_3^{(1)} = C_3^{(1)}, \quad D_2\overline{C}_3^{(2)} = C_3^{(2)}, \quad D_3\overline{C}_3^{(3)} = C_3^{(3)}.$$

We might have made use of $C_1^{(2)}$ and $C_1^{(3)}$, obtaining

$$
\begin{aligned}
C_2^{(2)} = {} & [(a_{35}a_{51} + a_{36}a_{61})y + (a_{35}a_{52} + a_{36}a_{62})z]\beta - [(a_{45}a_{51} + a_{46}a_{61})y \\
& + (a_{45}a_{52} + a_{46}a_{62})z]a \qquad (D_2\overline{C}_2^{(2)} = C_2^{(2)}),
\end{aligned} \tag{33}
$$

and then employed $C_1^{(1)}$ and $C_2^{(2)}$ in the same manner. But this would have resulted in $C_3^{(1)}$. Nor would anything new be obtained from $C_1^{(2)}$ and

$$
\begin{aligned}
C_2^{(3)} = {} & [(a_{51}a_{13} + a_{52}a_{23})\alpha + (a_{51}a_{14} + a_{52}a_{24})\beta]\zeta - [(a_{61}a_{13} + a_{62}a_{23})\alpha \\
& + (a_{61}a_{14} + a_{62}a_{24})\beta]\gamma \qquad (D_3\overline{C}_2^{(3)} = C_2^{(3)}).
\end{aligned} \tag{34}
$$

We have now found sixteen relative semicovariants, all of them homogeneous of the second degree in the dependent variables and all homogeneous in the coefficients a_{ij}. Since a transformation of the independent variable replaces a_{ij} with a_{ij}/ξ' and leaves $b_{\kappa\lambda}$, y, z, α, β, γ, ζ unchanged it follows that *these sixteen semicovariants are also relative covariants.* One of them, $C_0^{(1)}$, is in fact an absolute covariant since it does not contain a_{ij}.

Nine of the sixteen involve but two dependent variables each and these two correspond in each case to a pair of points determining a line of one of the three ruled surfaces R_{yz}, $R_{\alpha\beta}$, $R_{\gamma\zeta}$. We shall speak of these nine as *bivariants.* The remaining seven involve four dependent variables each and these four correspond in each case to two pairs of points determining lines from two of the three surfaces. These seven we shall term *quadrivariants.*

The significance of a quadratic covariant depends in general upon whether it can be factored. It is not difficult to show that *necessary and sufficient conditions for the factoring of quadrivariants are the vanishing of one or more of the quantities* A_1, A_2, A_3, B_1, B_2. Since we insist in this theory that these determinants shall be different from zero, we conclude that *the quadrivariants are irreducible.*

III. Geometric significance of invariants and covariants

We have already seen that the three lines l_{yz}, $l_{\alpha\beta}$, $l_{\gamma\zeta}$ determine a quadric, that the line of the regulus R_1 which passes through P_y cuts $l_{\alpha\beta}$ and $l_{\gamma\zeta}$ in the respective points $b_{13}\alpha+b_{14}\beta$, $b_{15}\gamma+b_{16}\zeta$, and that line which passes through P_z cuts $l_{\alpha\beta}$, $l_{\gamma\zeta}$ in the respective points $b_{23}\alpha+b_{24}\beta$, $b_{25}\gamma+b_{26}\zeta$. It follows that the line of R_1 which passes through the point P_ϕ of l_{yz} where $\phi=y+\mu z$, cuts $l_{\alpha\beta}$, $l_{\gamma\zeta}$ in the respective points

$$\theta = (b_{13}+\mu b_{23})\alpha + (b_{14}+\mu b_{24})\beta, \qquad \psi = (b_{15}+\mu b_{25})\gamma + (b_{16}+\mu b_{26})\zeta.$$

By differentiating and making use of (R) we have

$$\phi' = y' + \mu z' + \mu' z = a_{13}\alpha + a_{14}\beta + \mu(a_{23}\alpha + a_{24}\beta) + \mu' z,$$

so that

$$\phi' - \mu' z = (a_{13}+\mu a_{23})\alpha + (a_{14}+\mu a_{24})\beta. \tag{35}$$

Now $\phi'-\mu' z$ is a point on the tangent plane to R_{yz} at P_ϕ. It follows from (35) that this plane must cut $l_{\alpha\beta}$ in the point P_τ where

$$\tau = (a_{13}+\mu a_{23})\alpha + (a_{14}+\mu a_{24})\beta.$$

If that line of R_1 through P_ϕ is to be tangent to R_{yz} then it must lie in the tangent plane to R_{yz} at P_ϕ and hence must cut $l_{\alpha\beta}$ in the point in which this tangent plane cuts $l_{\alpha\beta}$. We have thus

$$\frac{a_{13}+\mu a_{23}}{b_{13}+\mu b_{23}} = \frac{a_{14}+\mu a_{24}}{b_{14}+\mu b_{24}},$$

or

$$(a_{23}b_{24}-a_{24}b_{23})\mu^2 + (a_{13}b_{24}-a_{14}b_{23}+a_{23}b_{14}-a_{24}b_{13})\mu + (a_{13}b_{14}-a_{14}b_{13}) = 0. \tag{36}$$

The two roots of (36) when substituted for μ in $\phi=y+\mu z$, give the expressions $\phi^{(1)}$, $\phi^{(2)}$, for the two points on l_{yz} at which the lines of R_1 are tangent to R_{yz}. The product $\phi^{(1)}\phi^{(2)}$ is precisely $K_1^{(1)}$.

The two points $P_{\phi^{(1)}}$, $P_{\phi^{(2)}}$ will coincide providing the discriminant of (36) vanishes. This discriminant, except for the factor B_1^2, is found to be $(I_1^{(1)})^2-4I_2^{(1)}$.

In view of the three-fold symmetry of our configuration it is unnecessary to repeat the above argument for the lines $l_{\alpha\beta}$, $l_{\gamma\zeta}$. By permuting the letters once we find that $K_1^{(1)}$ is replaced, except for the factor B_1, by $M_1^{(1)}$ and $(I_1^{(1)})^2-4I_2^{(1)}$ by $(I_1^{(2)})^2-4I_2^{(2)}$, and permuting a second time replaces $M_1^{(1)}$ by $L_1^{(2)}$ and $(I_1^{(2)})^2-4I_2^{(2)}$ by $I_1^2-4I_2$.

Summing up, we find that *there are two points on each of the lines* l_{yz}, $l_{\alpha\beta}$, $l_{\gamma\zeta}$ *at which the lines of the regulus* R_1 *are tangent to the respective ruled surfaces* R_{yz}, $R_{\alpha\beta}$, $R_{\gamma\zeta}$. *These pairs of points are given by factoring the respective covariants* $K_1^{(1)}$, $M_1^{(1)}$, $L_1^{(2)}$. *The conditions that the point pairs consist of coincident points are the vanishing of the respective invariants* $(I_1^{(1)})^2-4I_2^{(1)}$, $(I_1^{(2)})^2-4I_2^{(2)}$, $I_1^2-4I_2$.

The intersector tangents to R_{yz} at $P_{\phi^{(1)}}$ and $P_{\phi^{(2)}}$ must coincide with the lines of R_1 through these points, for otherwise the tangent planes to R_{yz} at these two points would both contain $l_{\alpha\beta}$ and hence would coincide in the plane determined (under these conditions) by l_{yz} and $l_{\alpha\beta}$. Our theorem may therefore be restated in terms of coincidence of intersector tangents with rulings of R_1.

The two points $P_{\theta^{(1)}}$ and $P_{\theta^{(2)}}$ in which $l_{\alpha\beta}$ is cut by the lines of R_1 through $P_{\phi^{(1)}}$ and $P_{\phi^{(2)}}$, are given by the expressions

$$\theta^{(1)} = (b_{13} + \mu_1 b_{23})\alpha + (b_{14} + \mu_1 b_{24})\beta, \quad \theta^{(2)} = (b_{13} + \mu_2 b_{23})\alpha + (b_{14} + \mu_2 b_{24})\beta.$$

If we form the product $\theta^{(1)}\theta^{(2)}$ and in it replace $\mu_1+\mu_2$ and $\mu_1\mu_2$ by their values from (36) we shall find that, except for the factor $B_1/(a_{24}b_{23}-a_{23}b_{24})$, the product reduces to the relative covariant $K_1^{(2)}$. Without further discussion we remark that similar interpretations attach to the covariants $M_1^{(2)}$ and $L_1^{(1)}$.

Differentiating the expression τ and making use of equations (R) we find

$$\tau' = (a_{13}' + \mu a_{23}' + \mu' a_{23})\alpha + (a_{14}' + \mu a_{24}' + \mu' a_{24})\beta + [a_{13}a_{35} + a_{14}a_{45} + \mu(a_{23}a_{35} + a_{24}a_{45})]\gamma + [a_{13}a_{36} + a_{14}a_{46} + \mu(a_{23}a_{36} + a_{24}a_{46})]\zeta.$$

Reasoning as before we find for the point in which the intersector tangent at P_τ cuts $l_{\gamma\zeta}$ the expression

$$\eta = [a_{13}a_{35} + a_{14}a_{45} + \mu(a_{23}a_{35} + a_{24}a_{45})]\gamma + [a_{13}a_{36} + a_{14}a_{46} + \mu(a_{23}a_{36} + a_{24}a_{46})]\zeta. \tag{37}$$

Let us suppose that the line $l_{\phi\tau}$ is tangent to $R_{\alpha\beta}$ at P_τ, so that it lies in the plane p tangent to $R_{\alpha\beta}$ at P_τ. The line $l_{\tau\eta}$ lies in this plane so that p cuts $l_{\gamma\zeta}$ in P_η. Since both P_ϕ and P_η lie in p, the line $l_{\phi\eta}$ must cut all three lines l_{yz}, $l_{\alpha\beta}$, $l_{\gamma\zeta}$. It is therefore that line of R_1 which passes through P_ϕ. It follows that P_η and P_ψ coincide and hence that

$$\frac{a_{13}a_{35} + a_{14}a_{45} + \mu(a_{23}a_{35} + a_{24}a_{45})}{a_{13}a_{36} + a_{14}a_{46} + \mu(a_{23}a_{36} + a_{24}a_{46})} = \frac{b_{15} + \mu b_{25}}{b_{16} + \mu b_{26}}.$$

This reduces to

$$L\mu^2 + M\mu + N = 0, \tag{38}$$

where

$$L = b_{26}(a_{23}a_{35} + a_{24}a_{45}) - b_{25}(a_{23}a_{36} + a_{24}a_{46}),$$
$$M = b_{26}(a_{13}a_{35} + a_{14}a_{45}) - b_{25}(a_{13}a_{36} + a_{14}a_{46}) + b_{16}(a_{23}a_{35} + a_{24}a_{45}) - b_{15}(a_{23}a_{36} + a_{24}a_{46}),$$
$$N = b_{16}(a_{13}a_{35} + a_{14}a_{45}) - b_{15}(a_{13}a_{36} + a_{14}a_{46}).$$

The two points on l_{yz} which correspond to the two roots of equation (38) are found to be given by the factors of the expression

$$Ly^2 - Myz + Nz^2$$

and this expression is the relative covariant obtained from $C_0^{(2)}$ and $C_2^{(1)}$ by the determinant process. *The factors of this covariant thus give the points on l_{yz} at which intersector tangents are also tangent to $R_{\alpha\beta}$.* These two points coincide providing $M^2-4LN=0$ and this expression reduces to the relative invariant

$$B_2^2\left[(J_2^{(2)})^2 - 4I_2^{(1)}I_2^{(2)}\right] = 0.$$

Corresponding pairs of points on $l_{\alpha\beta}$ and $l_{\gamma\zeta}$ are given by the factors of the two covariants obtained from the one above by permuting the letters and coincidence is conditioned upon the vanishing of the two invariants $J_2^2-4I_2^{(2)}I_2$ and $(J_2^{(1)})^2-4I_2I_2^{(1)}$.

Four skew lines in space have two, and only two, straight line intersectors. The two lines of R_1 which are tangent to $R_{yz}(R_{\alpha\beta}, R_{\gamma\zeta})$ furnish an illustration of this theorem. The two points in which they cut $l_{yz}(l_{\alpha\beta}, l_{\gamma\zeta})$ constitute in their entirety a curve of two branches on $R_{yz}(R_{\alpha\beta}, R_{\gamma\zeta})$. The two intersector tangents of R_{yz} $(R_{\alpha\beta}, R_{\gamma\zeta})$ which are also tangent to $R_{\alpha\beta}(R_{\gamma\zeta}, R_{yz})$ furnish a second illustration. The two points on $l_{yz}(l_{\alpha\beta}, l_{\gamma\zeta})$ at which these intersector tangents occur generate a curve of two branches on $R_{yz}(R_{\alpha\beta}, R_{\gamma\zeta})$. If we think of the asymptotic tangents to $R_{yz}(R_{\alpha\beta}, R_{\gamma\zeta})$ at points of $l_{yz}(l_{\alpha\beta}, l_{\gamma\zeta})$, there are evidently two which intersect $l_{\alpha\beta}(l_{\gamma\zeta}, l_{yz})$, and these two points at which they occur constitute for all the lines of $R_{yz}(R_{\alpha\beta}, R_{\gamma\zeta})$ a curve of two branches. Invariants and covariants connected with such curves would contain first derivatives of the coefficients.

For each of the three ruled surfaces R_{yz}, $R_{\alpha\beta}$, $R_{\gamma\zeta}$ we have thus three different curves, each consisting of two branches. Lane has discussed the second and third of these curves for a configuration of two ruled surfaces but without showing their relation to the invariants and covariants of the defining system of differential equations.*

* *Ruled surfaces in correspondence.*

Starting at any point P_ϕ whatever (designated hereafter as P_{ϕ_0}) on l_{yz}, we have progressed by means of intersector tangents through a point P_τ on $l_{\alpha\beta}$ to a point P_η on $l_{\gamma\delta}$. There is of course an intersector tangent to $R_{\gamma\delta}$ at P_η. By a repetition of the process of reasoning already twice employed we find an expression for the point P_{ϕ_1} in which this intersector tangent cuts l_{yz}. It is

$$\phi_1 = (A + \mu B)y + (C + \mu D)z, \tag{39}$$

where

$$\begin{aligned} A &= a_{13}a_{35}a_{51} + a_{13}a_{36}a_{61} + a_{14}a_{45}a_{51} + a_{14}a_{46}a_{61}, \\ B &= a_{23}a_{35}a_{51} + a_{23}a_{36}a_{61} + a_{24}a_{45}a_{51} + a_{24}a_{46}a_{61}, \\ C &= a_{13}a_{35}a_{52} + a_{13}a_{36}a_{62} + a_{14}a_{45}a_{52} + a_{14}a_{46}a_{62}, \\ D &= a_{23}a_{35}a_{52} + a_{23}a_{36}a_{62} + a_{24}a_{45}a_{52} + a_{24}a_{46}a_{62}. \end{aligned} \tag{40}$$

Now P_{ϕ_1} will coincide with P_{ϕ_0} if and only if, the coefficients of y and z in ϕ_1 and ϕ_0 are proportional, that is if

$$B\mu^2 + (A - D)\mu - C = 0. \tag{41}$$

Solving for μ and introducing the values so obtained into the expression for ϕ_0 we have

$$\begin{aligned} 2B\phi_0^{(1)} &= 2By + [D - A + ((A - D)^2 + 4BC)^{1/2}]z, \\ 2B\phi_0^{(2)} &= 2By + [D - A - ((A - D)^2 + 4BC)^{1/2}]z, \end{aligned}$$

or

$$B\phi_0^{(1)}\phi_0^{(2)} = By^2 + (D - A)yz - Cz^2.$$

But the right member of this equation is precisely the covariant $C_3^{(1)}$. It is not difficult to show that if we start with $l_{\alpha\beta}$ rather than l_{yz} we will obtain two points on $l_{\alpha\beta}$ whose expressions will be factors of the covariant $C_3^{(2)}$ and similarly two corresponding points on $l_{\gamma\delta}$ whose expressions will be factors of $C_3^{(3)}$.

Let us speak of four points related to each other as are P_{ϕ_0}, P_τ, P_η, P_{ϕ_1} as *an intersector sequence of order one, open if* $P_{\phi_0} \neq P_{\phi_1}$, *closed if* $P_{\phi_0} = P_{\phi_1}$. Then we have proved that *on each of the lines* l_{yz}, $l_{\alpha\beta}$, $l_{\gamma\delta}$, *there are two points for which the corresponding intersector sequences of order one are closed and these pairs of points are given respectively by factoring the three covariants* $C_3^{(1)}$, $C_3^{(2)}$, $C_3^{(3)}$. The sequences must be the same for all three pairs of points for otherwise there would be more than one such pair on each line.

We will suppose now that $P_{\phi_0} \neq P_{\phi_1}$ and let P_{ϕ_2} be that point of l_{yz} obtained from P_{ϕ_1} just as P_{ϕ_1} is obtained from P_{ϕ_0}. The set of points beginning with P_{ϕ_0} and ending with P_{ϕ_2}, seven in all, we shall speak of as *an intersector*

sequence of order two, closed if $P_{\phi_0}=P_{\phi_2}$, *open if* $P_{\phi_0}\neq P_{\phi_2}$. We find for the expression for this point

$$\phi_2 = (A^{(1)} + \mu B^{(1)})y + (C^{(1)} + \mu D^{(1)})z,$$

where

$$(42)\quad A^{(1)} = A^2 + BC,\ B^{(1)} = B(A + D),\ C^{(1)} = C(A + D),\ D^{(1)} = BC + D^2.$$

The point P_{ϕ_2} will coincide with the point P_{ϕ_0} if

$$B^{(1)}\mu^2 + (A^{(1)} - D^{(1)})\mu - C^{(1)} = 0.$$

But this reduces to

$$(A + D)[B\mu^2 + (A - D)\mu - C] = 0.$$

The vanishing of the second factor gives those values of μ to which correspond the values $\phi_0^{(1)}$, $\phi_0^{(2)}$ of ϕ_0. The presence of this factor was to be expected since a closed sequence of order one, twice followed, constitutes a special case of a closed sequence of order two. It is the vanishing of the factor $(A+D)$ that is significant. For by referring to (35) it is seen that if $A+D=0$, $P_{\phi_0}=P_{\phi_2}$ for every point whatever of l_{yz}. Now

$$A + D = a_{13}(a_{35}a_{51} + a_{36}a_{61}) + a_{14}(a_{45}a_{51} + a_{46}a_{61}) + a_{23}(a_{35}a_{52} + a_{36}a_{62}) + a_{24}(a_{45}a_{52} + a_{46}a_{62}) = J_3.$$

We have thus proved that *if the invariant* $J_3=0$, *the intersector sequences of the second order for all points of* l_{yz} *are closed.* We remark that *the points of* $l_{\alpha\beta}$, $l_{\gamma\xi}$ *also have this same property when* $J_3=0$.

Before investigating intersector sequences of higher order we note by (41) that the two points of l_{yz} whose intersector sequences of order one are closed will coincide if

$$(A - D)^2 + 4BC = 0.$$

This expression is a relative invariant $K=J_3^2-4I_6$. If $J_3=0$ then $K\neq 0$, since $I_6\neq 0$. If the intersector sequences of order two for all the points of l_{yz} are closed then the two distinct points of l_{yz} whose sequences of the second order consist of a sequence of the first order once repeated, are given by the roots of the equation

$$B\mu^2 + 2A\mu - C = 0.$$

Assuming now that P_{ϕ_0}, P_{ϕ_1}, P_{ϕ_2} are in general distinct, we seek to find the expression for the point P_{ϕ_3} which, together with P_{ϕ_0}, P_{ϕ_1}, P_{ϕ_2} and the

intervening points on $l_{\alpha\beta}$, $l_{\gamma\xi}$, constitutes a sequence of order three. Proceeding as before we find

$$\phi_3 = (A^{(2)} + \mu B^{(2)})y + (C^{(2)} + \mu D^{(2)})z,$$

where

$$\begin{aligned} A^{(2)} &= (A^2 + BC)^2 + BC(A + D)^2, \\ B^{(2)} &= B(A + D)(A^2 + 2BC + D^2), \\ C^{(2)} &= C(A + D)(A^2 + 2BC + D^2), \\ D^{(2)} &= BC(A + D)^2 + (BC + D^2)^2, \end{aligned} \tag{43}$$

and this point coincides with P_{ϕ_1} if

$$B^{(2)}\mu^2 + (A^{(2)} - D^{(2)})\mu - C^{(2)} = 0,$$

that is, if

$$(A + D)(A^2 + 2BC + D^2)[B\mu^2 + (A - D)\mu - C] = 0.$$

The significant factor this time is $A^2+2BC+D^2$. It proves to be the relative invariant $K_3=J_3^2-2I_6$. If $A^2+2BC+D^2=0$ it is seen by (43) that $\phi_0=\phi_3$ for all values of μ. In this case, as before, $A+D=J_3$ cannot vanish when $J_3^2-2I_6=0$.

For intersector sequences of order four we find for P_{ϕ_4} the expression

$$\phi_4 = (A^{(3)} + \mu B^{(3)})y + (C^{(3)} + \mu D^{(3)})z,$$

where

$$\begin{aligned} A^{(3)} &= [(A^2 + BC)^2 + BC(A + D)^2]^2 + BC(A + D)^2(A^2 + 2BC + D^2)^2, \\ B^{(3)} &= B(A + D)(A^2 + 2BC + D^2)[(A^2 + BC)^2 + 2BC(A + D)^2 + (BC + D^2)^2], \\ C^{(3)} &= C(A + D)(A^2 + 2BC + D^2)[(A^2 + BC)^2 + 2BC(A + D)^2 + (BC + D^2)^2], \\ D^{(3)} &= BC(A + D)^2(A^2 + 2BC + D^2)^2 + [BC(A + D)^2 + (BC + D^2)^2]^2. \end{aligned}$$

$P_{\phi_0}=P_{\phi_4}$ if

$$\begin{aligned} (A + D)(A^2 + 2BC + D^2)[(A^2 + BC)^2 + 2BC(A + D)^2 \\ + (BC + D^2)^2][B\mu^2 + (A - D)\mu - C] = 0. \end{aligned}$$

The significant factor $(A^2+BC)^2+2BC(A+D)^2+(BC+D^2)^2$ reduces to the relative invariant $K_4=J_3^4-4J_3^2I_6-2I_6^2=K_3^2-2I_6^2$. As before K_4 cannot vanish when $K_3=0$.

There is no need to repeat the analysis for sequences of orders 5, 6, $\cdots$. The reasoning is general. For intersector sequences of order n the necessary and sufficient condition for closure is the vanishing of the relative invariant

$$K_n = K_{n-1}^2 - 2I_6^{2^{n-3}} \qquad (n = 3,4, \cdots),$$

where

$$K_2 = J_3.$$

The vanishing of any K_n precludes the vanishing of all other K's. K_n is of weight $3 \cdot 2^{n-2}$ and of the same degree in the a_{ij}. To recapitulate: *to each point of the line l_{yz} of R_{yz} ($l_{\alpha\beta}$ of $R_{\alpha\beta}$, $l_{\gamma\zeta}$ of $R_{\gamma\zeta}$) there corresponds an intersector sequence of order n ($n=2, 3, \cdots$). The necessary and sufficient condition for closure of all such sequences of order n is the vanishing of an invariant of weight and degree $3 \cdot 2^{n-2}$. The possession of the closure property by the sequences of any given order precludes the possession of this property by the sequences of all other orders. To only two points of $l_{yz}(l_{\alpha\beta}, l_{\gamma\zeta})$ correspond closed sequences of order one.*

We have now obtained geometrical interpretations of the ten relative invariants or combinations of them, and of the nine bivariants. Many other interesting problems present themselves for investigation: the properties of the two congruences made up, in the one case of the one-parameter family of reguli R_1 and in the other of the reguli R_2 to which second congruence the surfaces R_{yz}, $R_{\alpha\beta}$, $R_{\gamma\zeta}$ belong; the relations of the invariants and covariants of the separate surfaces R_{yz}, $R_{\alpha\beta}$, $R_{\gamma\zeta}$ to those of the whole configuration; the questions of independence of the invariants and covariants already found, the maximum number of invariants (covariants) of a given kind, the determination of a fundamental system of invariants (covariants) in the sense of a set in terms of which, and their derivatives, all invariants (covariants) can be expressed. These considerations must be left for treatment in a subsequent paper.

UNIVERSITY OF WASHINGTON,
SEATTLE, WASH.

CERTAIN UNIFORM FUNCTIONS OF RATIONAL FUNCTIONS*

BY

EMORY P. STARKE

I. Introduction

A natural generalization of the notion of an even function (i.e., a uniform function of z^2) is that of a uniform function of any rational function of z. In this paper, we examine the following classes of functions and determine which functions in each class are uniform functions of non-linear rational functions:

(1) periodic meromorphic functions;

(2) Poincaré's functions with rational multiplication theorems.†

Among the periodic functions, we shall find in §III that only the following possess the desired property:

(i) Meromorphic functions of $\cos(\lambda z+\mu)$. These are uniform functions of $(z+\mu/\lambda)^2$.

(ii) Rational functions of $\wp(z+\mu)$. These are uniform functions of $(z+\mu)^2$.

(iii) Rational functions of $\wp^2(z+\mu)$ in the lemniscatic case $(g_3=0)$. These are functions of $(z+\mu)^4$.

(iv) Rational functions of $\wp'(z+\mu)$ or of $\wp^3(z+\mu)$ in the equianharmonic case $(g_2=0)$. These are functions of $(z+\mu)^3$ or of $(z+\mu)^6$ respectively.

In the above λ and μ are arbitrary constants.

For the Poincaré functions (§IV), our results are as follows:

(a) If in any Poincaré function we replace z with z^q, where q is an integer, we shall have another Poincaré function which is a uniform function of z^q.

(b) Every Poincaré function which is a uniform function of a non-linear rational function and which is not uniform in any z^q, is a periodic function. Using the results of a paper by Ritt, we see that such functions are necessarily linear functions of one of the functions $\cos(\lambda z+\mu)$, $\wp(z+\mu)$; in the lemniscatic case, $\wp^2(z+\mu)$; in the equianharmonic case, $\wp'(z+\mu)$ and $\wp^3(z+\mu)$. Here μ is restricted to certain values.

* Presented to the Society, May 1, 1926; received by the editors in May, 1926.

† These are described in detail below (§IV).

II. Uniform functions of rational functions: some general results

Let $f(z)$ be a non-constant, meromorphic function satisfying a relation

$$f(z) = g[\sigma(z)], \tag{1}$$

with $\sigma(z)$ rational and not linear, and with $g(z)$ uniform. Without loss of generality, we may require $g(z)$ to be also meromorphic; for, since $f(z)$ is meromorphic, $g(z)$ cannot have other essential singularities than $\sigma(\infty)$. We may suppose, replacing $\sigma(z)$ by a linear function of itself if necessary, that $\sigma(\infty)=\infty$. Having now no finite essential singularity, $g(z)$ is meromorphic.

$\sigma(z)$ then is a polynomial; for, if it had a pole at a finite point, this point would be an essential singularity for the meromorphic $f(z)$. We may now put (multiplying $\sigma(z)$ by a constant if necessary to give unity for the leading coefficient)

$$\sigma(z) = z^n + b_1 z^{n-1} + \cdots + b_n. \tag{2}$$

The n branches of the inverse of $u=\sigma(z)$ have, for the neighborhood of infinity, the single expansion

$$z = \sigma^{-1}(u) = u^{1/n} + c_0 + c_1 u^{-1/n} + \cdots. \tag{3}$$

Consider one of the branches z in (3), and the branch z' by which it is replaced when u describes a single circuit about a large circle in the positive direction. We are going to express z' in terms of z.

For a single circuit of u, $u^{1/n}$ goes over into $\epsilon u^{1/n}$ where $\epsilon=e^{2\pi i/n}$. If then in (3) we replace $u^{1/n}$ by $\epsilon u^{1/n}$ and substitute for $u^{1/n}$ that branch of $[\sigma(z)]^{1/n}$ which is used in the calculation of the original branch z in (3), we shall have z' expressed as a function of z. Writing $z'=\alpha(z)$, we find thus, for $|z|$ sufficiently large,

$$\alpha(z) = \epsilon z + a_0 + \sum_{n=1}^{\infty} a_n z^{-n}. \tag{4}$$

The principal part of the proofs in both of the following sections will be to show that in the cases there treated $\alpha(z)$ is linear. Let us assume then that $\alpha(z)$ is not linear, and fix our attention on the first non-vanishing coefficient of the summation in (4). If its subscript is p, equation (4) now takes the form

$$\alpha(z) = \epsilon z + a_0 + a_p z^{-p} + \cdots, \qquad a_p \neq 0, \qquad p \geqq 1. \tag{4'}$$

We shall use also the inverse of $\alpha(z)$, which, like $\alpha(z)$, is analytic for $|z|$ large. Its development for the neighborhood of infinity is, if $p>1$,

$$\alpha^{-1}(z) = \frac{z-a_0}{\epsilon} - \frac{\epsilon^{p-1}a_p}{z^p} - \frac{\epsilon^{p-1}(pa_0a_p + \epsilon a_{p+1})}{z^{p+1}} - \frac{\epsilon^{p-1}[\frac{1}{2}(p^2+p)a_0^2a_p + (p+1)\epsilon a_0a_{p+1} + \epsilon^2 a_{p+2}]}{z^{p+2}} + \cdots. \tag{5}$$

If $p=1$, the term $+\epsilon a_1^2$ should be added to the expression in brackets. But this peculiarity of the case $p=1$ does not carry over into the calculations of the following sections; we secure the same equations for $p>1$ and for $p=1$.

Of course, for $|z|$ large, we have

$$\sigma(z) = \sigma[\alpha(z)] = \sigma[\alpha^{-1}(z)], \tag{6}$$

and thus

$$f(z) = f[\alpha(z)] = f[\alpha^{-1}(z)].$$

III. Periodic functions

In addition to the hypotheses of the foregoing section, let $f(z)$ be periodic, either simply or doubly, and let h be any period of $f(z)$. We take any finite point z, of large modulus, and proceed to find certain other points at which $f(z)$ assumes the same value as at z. We form the expressions

$$z_1 = \alpha(z) + h = \epsilon z + a_0 + h + a_p z^{-p} + a_{p+1}z^{-p-1} + a_{p+2}z^{-p-2} + \cdots, \tag{7a}$$

$$z_2 = \alpha^{-1}(z_1) + h = z + h + \frac{h}{\epsilon} + \frac{pa_ph}{\epsilon^2 z^{p+1}} + \frac{2\epsilon(p+1)a_{p+1}h - p(p+1)a_ph^2}{2\epsilon^3 z^{p+2}} + \cdots, \tag{7b}$$

$$z_3 = \alpha(z_2) - h = \epsilon z + a_0 + \epsilon h + \frac{a_p}{z^p} + \frac{a_{p+1}}{z^{p+1}} + \frac{a_{p+2}}{z^{p+2}} - \frac{pa_ph}{z^{p+1}} + \frac{(p+1)[pa_ph^2(2+\epsilon) - 2\epsilon h a_{p+1}]}{2\epsilon z^{p+2}} + \cdots. \tag{7c}$$

Putting $\phi(z) = \alpha^{-1}(z_3) - h$, we find

$$\phi(z) = z + \frac{p(p+1)h^2a_p}{\epsilon^2 z^{p+2}} + \cdots. \tag{8}$$

In view of (6) and of the relation $f(z \pm h) = f(z)$, we have

$$f[\phi(z)] = f(z_j) = f(z) \qquad (j = 1,2,3). \tag{9}$$

If $|z|$ is sufficiently large, there will be a z_1, a z_2, a z_3, and a $\phi(z)$ which will satisfy (9).

To fix our ideas, suppose that $|z+h| \geqq |z|$. Consider the points $z+jh$ $(j=1, 2, \cdots)$. (If $|z+h| < |z|$, take the points $z-jh$, instead.) For all these points, the relation (9) holds. If now, according to (8), we form the expression,

$$\phi(z + jh) - jh = z + \frac{p(p+1)h^2 a_p}{\epsilon^2 (z+jh)^{p+2}} + \cdots,$$

we see that the left-hand member gives, as j assumes all positive integral values, a set of points of which infinitely many are in every neighborhood of z, but for each of which the function $f(z)$ takes the same value as at z. Since z is not an essential singular point of the meromorphic function $f(z)$, we have a contradiction. The assumption that $\phi(z)$ is not linear, which we made in (4′), is therefore untenable. We have then

$$\alpha(z) = \epsilon z + a_0. \tag{10}$$

If we differentiate the relation

$$\sigma(z) = \sigma[\alpha(z)] = \sigma(\epsilon z + a_0)$$

s times, we obtain the equation

$$\sigma^{(s)}(z) = \epsilon^s \sigma^{(s)}(\epsilon z + a_0).$$

Noting that $\epsilon[a_0/(1-\epsilon)] + a_0 = a_0/(1-\epsilon)$, and that $\epsilon^s \neq 1$ for $s=1, 2, \cdots, n-1$, we see that the first $n-1$ derivatives of $\sigma(z)$ must vanish for the value $z = a_0/(1-\epsilon)$. The polynomial $\sigma(z)$ has therefore the form

$$\sigma(z) = \left[z - \frac{a_0}{1-\epsilon}\right]^n + k,$$

in which k is a constant. Since we may still add to $\sigma(z)$ an arbitrary constant if necessary, we put

$$\sigma(z) = (z-c)^n, \qquad c = a_0/(1-\epsilon). \tag{11}$$

Referring to equation (7b) and remembering that now $a_p = 0$, $a_{p+1} = 0$, $\cdots$, we see that h/ϵ is a period of $f(z)$. As h is any period, it follows that h multiplied by any power of ϵ will be a period. Since, further, $\sigma(z)$ is not linear, we must have $n > 1$ and $\epsilon \neq 1$.

If $f(z)$ is simply periodic, with periods h and h/ϵ, evidently ϵ must be -1 and n must be 2. We have

$$\alpha(z) = 2c - z, \qquad \sigma(z) = (z - c)^2,$$
$$f(z) = f(2c - z) = g[(z - c)^2].$$

If we write $c+hu/2\pi$ for z, $c'h/2\pi$ for c and $F(2\pi z/h)$ for $f(z)$, the relation $f(z)=f(2c-z)$ becomes

$$F(c' + u) = F(c' - u),$$

where $F(u)$ has the period 2π. Being also even with respect to u, $F(c'+u)$ must be a uniform function of $\cos u$. Replacing the original letters, we have $f(z)$ expressed as a meromorphic function of $\cos 2\pi(z-c)/h$. With a slight change of notation, this gives

$$f(z) = \psi[\cos(\lambda z + \mu)] = g[(z + \mu/\lambda)^2], \tag{12}$$

with $\psi(z)$ and $g(z)$ meromorphic. Any meromorphic $\psi(z)$ will give an $f(z)$ which is uniform in $(z+\mu/\lambda)^2$.

If now $f(z)$ is doubly periodic, it is of course a rational function of $\wp(z)$ and $\wp'(z)$. It may be shown that n must be 2 unless $g_3=0$ or $g_2=0$. If $g_3=0$, n may be 2 or 4, while if $g_2=0$, n may be 2, 3, or 6.*

First consider the case $n=2$. We have $f(z)=f(2c-z)$. If we set $z=u+c$, this becomes $f(c+u)=f(c-u)$. If the periods of $f(z)$ are $2\omega_1$ and $2\omega_3$, $f(c+u)$, being even with respect to u, is a rational function of $\wp(u/\omega_1, \omega_3)$ alone. We then have

$$f(z) = \psi[\wp(z - c)] = g[(z - c)^2], \tag{13}$$

with $\psi(z)$ rational.

If now $n=3$, we have $f(z)=f[\alpha(z)]$ or $f(z)=f[\epsilon z+(1-\epsilon)c]$. Setting $z=u+c$, we find that $f(u+c)=f(\epsilon u+c)$. Now in the equianharmonic case, $\wp'(\epsilon u)=\wp'(u)$. Then both $f(u+c)$ and $\wp'(u)$ remain unchanged if u is replaced by ϵu. Hence in the development of either function at the origin, the exponents of u are all multiples of 3. Since both functions are uniform functions of u^3, while $\wp'(u)$ is of the third order, we see that $f(u+c)$ is a rational function of $\wp'(u)$. We have then

$$f(z) = \psi[\wp'(z - c)] = g[(z - c)^3], \tag{14}$$

with $\psi(z)$ again rational.

---

* See Ritt, *Periodic functions with a multiplication theorem*, these Transactions, vol. 23 (1922), p. 20.

We handle the remaining cases in precisely the same manner, obtaining the following results: For $n=4$ (lemniscatic case),

$$f(z) = \psi[\wp^2(z-c)] = g[(z-c)^4]. \tag{15}$$

For $n=6$ (equianharmonic case),

$$f(z) = \psi[\wp^3(z-c)] = g[(z-c)^6]. \tag{16}$$

In (15) and (16), $\psi(z)$ is rational.

IV. The Poincaré functions

Let $f(z)$ be a non-constant, meromorphic function of z such that $f(mz)$ is a rational function of $f(z)$,

$$f(mz) = R[f(z)], \tag{17}$$

where m is any constant of modulus greater than unity. Functions $f(z)$ of this type formed the object of an investigation by Poincaré.* We shall investigate the circumstances under which $f(z)$ can be put in the form,

$$f(z) = g[\sigma(z)], \tag{18}$$

in which $g(z)$ is meromorphic and $\sigma(z)$ is a non-linear polynomial as in §II.

We indicate here a few properties of the functions $f(z)$ and $R(z)$ which will be useful in the following discussion.†

(i) We may assume, without loss of generality, upon replacing each of the functions $f(z)$ and $R(z)$, if necessary, by a suitable linear function of itself, that $f(0)=0$ and $R(0)=0$.

(ii) If $R(z)$ is linear, $f(z)$ is a linear function of some z^q (q an integer). $f(z)$ has then a zero of order q at the origin.

(iii) If $f(z)$ has a zero of order q at the origin, $f(z)$ is of the form $g(z^q)$, in which $g(z)$ is a Poincaré function with a simple zero at the origin, and satisfies the relation

$$g(Mz) = R[g(z)], \qquad M = m^q = R'(0).$$

(iv) If $g(z)$ is any Poincaré function, the function $f(z)=g(z^q)$ clearly satisfies both relations (17) and (18). These functions, $g(z^q)$, are among the functions we seek. This gives result (a) of the introduction.

---

* H. Poincaré, *Sur une classe nouvelle de transcendantes uniformes*, Journal de Mathématiques, (4), vol. 55 (1890).

† Noted by Ritt, *Transcendental transcendency of certain functions of Poincaré*, Mathematische Annalen, vol. 95 (1926).

(v) If $R(z)$ is not linear, $f(z)$ assumes the value zero at an infinite number of points.

In what follows, we set aside the functions described in (iii) and (iv) above. There remain to be considered only Poincaré functions, $f(z)$, for which we have $f(0)=0$, $f'(0)\neq 0$, with $R(z)$ a non-linear rational function such that $R(0)=0$, $R'(0)=m$.

From equation (6), §II, we have, for $|z|$ sufficiently large,

$$f[\alpha(z)] = f(z), \qquad f[\alpha(mz)] = f(mz). \tag{19}$$

Here again, we desire to show that $\alpha(z)$ is linear. Let us then assume, as in (4′), that it is not, but contains a term $a_p z^{-p}\neq 0$, $p\geqq 1$. Using (17), we may rewrite the second of the relations (19) thus:

$$R\{f[\beta(z)]\} = R\left\{f\left[\frac{\alpha(mz)}{m}\right]\right\} = R\{f(z)\},$$

in which, for simplicity of notation, we have put

$$\beta(z) = [\alpha(mz)]/m.$$

The calculation of $\beta(z)$ and its inverse in terms of the coefficients of (4′) and (5) gives, for $|z|$ large,

$$\beta(z) = \epsilon z + \frac{a_0}{m} + \frac{a_p}{m^{p+1}z^p} + \frac{a_{p+1}}{m^{p+2}z^{p+1}} + \cdots, \tag{20}$$

$$\beta^{-1}(z) = \frac{z}{\epsilon} - \frac{a_0}{m\epsilon} - \frac{\epsilon^{p-1}a_p}{m^{p+1}z^p} - \frac{\epsilon^{p-1}(pa_0a_p + \epsilon a_{p+1})}{m^{p+2}z^{p+1}} + \cdots. \tag{21}$$

We take now a point z of sufficiently large modulus and seek, by combining $\alpha(z)$, $\beta(z)$ and their inverses, to obtain a point close to z, at which the function $R[f(z)]$ has the same value as at z. We have*

$$\alpha^2(z) = \epsilon^2 z + (\epsilon+1)a_0 + \frac{a_p(\epsilon^{p+1}+1)}{\epsilon^p z^p} + \frac{a_{p+1}(\epsilon^{p+2}+1) - pa_0a_p}{\epsilon^{p+1}z^{p+1}} + \cdots,$$

$$\beta^{-1}\alpha^2(z) = \epsilon z + a_0\left(\frac{m\epsilon+m-1}{m\epsilon}\right) + \frac{a_p}{z^p}\left(\frac{m^{p+1}\epsilon^{p+1}+m^{p+1}-1}{m^{p+1}\epsilon^{p+1}}\right)$$
$$+ \frac{1}{z^{p+1}}\left[a_{p+1}\left(\frac{m^{p+2}\epsilon^{p+2}+m^{p+2}-1}{m^{p+2}\epsilon^{p+2}}\right) - pa_0a_p\left(\frac{\epsilon m^{p+2}-\epsilon m-m+1}{m^{p+2}\epsilon^{p+3}}\right)\right] + \cdots,$$

* We abbreviate $\beta^{-1}\{\alpha[\alpha(z)]\}$ thus: $\beta^{-1}\alpha^2(z)$. Similarly for the other functions.

$$\alpha^{-1}\beta^{-1}\alpha^2(z) = z + \frac{a_0(m-1)}{m\epsilon^2} + \frac{a_p(m^{p+1}-1)}{m^{p+1}\epsilon^{p+2}z^p} + \frac{1}{z^{p+1}}\left[a_{p+1}\left\{\frac{(m^{p+2}-1)}{m^{p+2}\epsilon^{p+3}}\right\} + pa_0a_p\left\{\frac{(\epsilon^{p+1}m^{p+1}+1)(m-1)-\epsilon m(m^{p+1}-1)}{m^{p+2}\epsilon^{p+4}}\right\}\right] + \cdots.$$

Similarly, we have

$$\beta^{-1}\alpha^{-1}\beta^2(z) = z - \frac{a_0(m-1)}{m\epsilon^2} - \frac{a_p(m^{p+1}-1)}{m^{p+1}\epsilon^{p+2}z^p} - \frac{1}{z^{p+1}}\left[a_{p+1}\left\{\frac{(m^{p+2}-1)}{m^{p+2}\epsilon^{p+3}}\right\} + pa_0a_p\left\{\frac{(\epsilon^{p+1}+m^{p+1})(m-1)-\epsilon(m^{p+1}-1)}{m^{p+2}\epsilon^{p+4}}\right\}\right] + \cdots.$$

We now form the combination

$$\phi(z) = \beta^{-1}\alpha^{-1}\beta^2\alpha^{-1}\beta^{-1}\alpha^2(z).$$

This gives

$$\phi(z) = z + \frac{0}{z^p} + \frac{pa_0a_p(m-1)(m^{p+1}-1)(\epsilon^p-1)}{m^{p+2}\epsilon^{p+3}z^{p+1}} + \cdots. \tag{22}$$

Since we had

$$R\{f[\alpha(z)]\} = R\{f(z)\} = R\{f[\beta(z)]\}$$

and

$$R\{f[\alpha^{-1}(z)]\} = R\{f(z)\} = R\{f[\beta^{-1}(z)]\},$$

we shall have also

$$R\{f[\phi(z)]\} = R\{f(z)\}. \tag{23}$$

We put d_k for the coefficient of z^{-k} in (22), so that (22) becomes

$$\phi(z) = z + d_pz^{-p} + d_{p+1}z^{-p-1} + d_{p+2}z^{-p-2} + \cdots. \tag{24}$$

If $|z|$ is large enough there will be a point $\phi(z)$ given by (24) which will satisfy the relation (23). But we shall show that $\phi(z)$, as given by (24), approximates too closely to z for (23) to hold, unless $\phi(z)=z$.

Suppose that $\phi(z)$ is not identically z. Let γ be a point of large modulus for which we have $f(\gamma)=0$. Evidently, we shall have also $f(m^s\gamma)=0$, for any integer s.

We consider, in addition, the point $\zeta=\phi(m^s\gamma)/m^s$, for which we have the relation

$$R^{s+1}[f(\zeta)] = R[f\{\phi(m^s\gamma)\}] = R[f(m^s\gamma)] = R^{s+1}[f(\gamma)] = 0. \tag{25}$$

$R(z)$ has a development about the origin of the form

$$R(z) = mz + \cdots, \qquad |m| > 1. \tag{26}$$

There exists a positive number ρ such that, in (26), we have

$$|m^{1/2}z| < |R(z)| < |m^{3/2}z|, \quad \text{for} \quad 0 < |z| \leqq \rho. \tag{27}$$

Since $f(\gamma)=0$, we can develop $f(z)$, for the neighborhood of γ, in the form

$$f(z) = A_j(z-\gamma)^j + A_{j+1}(z-\gamma)^{j+1} + \cdots \qquad (j \geqq 1). \tag{28}$$

Using (24), we find

$$\zeta = \frac{\phi(m^s\gamma)}{m^s} = \gamma + \frac{d_p}{m^{s(p+1)}\gamma^p} + \frac{d_{p+1}}{m^{s(p+2)}\gamma^{p+1}} + \cdots. \tag{29}$$

For s large, $\zeta \neq \gamma$, so that $f(\zeta) \neq 0$.* For sufficiently large values of s, we may substitute the result of (29) into (28), and obtain

$$f(\zeta) = \frac{A_j d_p{}^j}{m^{js(p+1)}\gamma^{jp}} + \frac{jA_j d_p{}^{j-1}d_{p+1}}{m^{s(pj+j+1)}\gamma^{pj+1}} + \cdots. \tag{30}$$

The absolute value of the denominator of the first term in (30) is not less than $|m^{2s}|$. Hence, for s large, we have

$$|f(\zeta)| < \rho/|m^{3s/2}| < \rho. \tag{31}$$

From (27) and (31) we obtain, for s large,

$$|m^{1/2}f(\zeta)| < |R[f(\zeta)]| < |m^{3/2}f(\zeta)| < |\rho/m^{3(s-1)/2}|.$$

We may therefore put $R[f(\zeta)]$ for z in (27), and have

$$|m^{1/2}R[f(\zeta)]| < |R^2[f(\zeta)]| < |m^{3/2}R[f(\zeta)]|.$$

But this gives

$$|mf(\zeta)| < |R^2[f(\zeta)]| < |\rho/m^{3(s-2)/2}| < \rho.$$

Continuing in this way, we find

$$|m^{t/2}f(\zeta)| < |R^t[f(\zeta)]| < |\rho/m^{3(s-t)/2}| < \rho \qquad (t=1, 2, \cdots, s),$$

* It is understood in this that γ is first chosen and held fixed, and that sufficiently large values are then given to s.

and finally,

$$|m^{(s+1)/2}f(\zeta)| < |R^{s+1}[f(\zeta)]| < \rho|m|^{3/2}. \tag{32}$$

The contradiction between (25) and (32) shows that $\phi(z)$ must be identically z. The coefficients d in (24) are thus zero. In particular we have

$$d_{p+1} = pa_0a_p(m-1)(m^{p+1}-1)(\epsilon^p-1)/(m^{p+2}\epsilon^{p+3}) = 0.$$

Recalling that $p \geqq 1$, $|m|>1$ and that a_p is assumed to be different from zero, we have either $a_0=0$ or $\epsilon^p=1$.

We see that a_0 cannot be zero from the following considerations. Suppose $a_0=0$, and form from (4′) and (21) the function

$$\alpha\beta^{-1}(z) = z + \frac{(m^{p+1}-1)\epsilon^p a_p}{m^{p+1}z^p} + \cdots.$$

Now we have $R\{f[\alpha\beta^{-1}(z)]\} = R\{f(z)\}$. If we represent $\alpha\beta^{-1}(z)$ by $\phi(z)$, the discussion above from equations (23) to (32) will hold for this new $\phi(z)$. We are led to the result $a_p=0$, which is supposed not to be true. Hence $a_0 \neq 0$.

On the other hand, ϵ^p cannot be 1. To show this, we expand the polynomial $\sigma(z)=\sigma[\alpha(z)]$ in the form

$$\sigma(z) = \sigma(\epsilon z + a_0) + \sigma'(\epsilon z + a_0)\left[\frac{a_p}{z^p} + \frac{a_{p+1}}{z^{p+1}} + \cdots\right] + \frac{\sigma''(\epsilon z + a_0)}{2!}\left[\frac{a_p}{z^p} + \frac{a_{p+1}}{z^{p+1}} + \cdots\right]^2 + \cdots. \tag{33}$$

If now $p \geqq n$, there is in (33) only one term of degree $(n-1-p)$. Its coefficient, $n\epsilon^{n-1}a_p$, however, cannot be zero. Hence p must be less than n. But if $\epsilon^p=1$, since $\epsilon=e^{2\pi i/n}$, we must have p a multiple of n. It follows that $\epsilon^p \neq 1$.

The contradiction noted in equation (32) cannot be removed under our assumption of the existence of an $a_p \neq 0$. Hence $\alpha(z)$ is linear, and we have, as in §III, the relations

$$\alpha(z) = \epsilon z + a_0, \quad \sigma(z) = (z-c)^n, \quad a_0 = (1-\epsilon)c.$$

So then we have $f(z)=f(\epsilon z+a_0)$, and also

$$f(z) = R^s[f(z/m^s)] = R^s[f(\epsilon z/m^s + a_0)] = f(\epsilon z + m^s a_0).$$

We obtain thus the relation

$$f(\epsilon z + a_0) = f(\epsilon z + m^s a_0). \tag{34}$$

Here too, a_0 cannot be zero. For, if $f(z)=f(\epsilon z)$, it would follow that $f(z)$ is a meromorphic function of z^n, and $f'(0)$ is zero. But such functions have previously been set aside.

Equation (34) then recites the fact that $f(z)$ is a periodic function with $(m^s-1)a_0$, a period. This establishes result (b) of the introduction: every Poincaré function which is not uniform in any z^q, and which satisfies a relation (18), is a periodic function.

Periodic functions which satisfy (18) are listed in §III. Periodic functions which satisfy (17) are given in a paper by Ritt.* The comparison of these results shows that periodic Poincaré functions which are uniform functions of non-linear rational functions are linear functions of one of the functions $\cos(\lambda z+\mu)$, $\wp(z+\mu)$; $\wp^2(z+\mu)$ if $g_3=0$; $\wp'(z+\mu)$ and $\wp^3(z+\mu)$ if $g_2=0$. Here μ is an appropriate fraction of a period.

---

* Ritt, *Periodic functions with a multiplication theorem*, these Transactions, vol. 23 (1922), pp. 16–25.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

ON REJECTION TO INFINITY AND EXTERIOR MOTION IN THE RESTRICTED PROBLEM OF THREE BODIES*

BY

BERNARD OSGOOD KOOPMAN

INTRODUCTION

The cases of the restricted problem of three bodies which have hitherto been considered in detail deal with motion of the particle in the interior of a closed oval of zero velocity, or in the neighborhood of points of equilibrium or periodic orbits. In none of these cases does the question of rejection of the infinitesimal body to infinity play any rôle. It is to the investigation of this phenomenon, and of types of motion in which its possibility is an essential feature, that the present paper is devoted. Attention is restricted to the case of motion in a plane.

Chapter I consists of a qualitative study of the individual half orbits which are described by the particle as it goes to or comes from infinity. The methods are altogether elementary. The first salient result is the proof of the existence of such half orbits, with a practical criterion for finding them. The second is the theorem that every half orbit which extends to infinity resembles a hyperbolic or parabolic orbit in this: that it has an asymptote (in the fixed space), and the particle has a limiting velocity. In a special case, resembling the parabolic, this velocity is zero and the asymptote is at infinity (projectively speaking).

In Chapter II, we take up the investigation of infinity from the point of view of a singularity of the differential equations of motion. It is shown that by means of a rational substitution this singularity can be reduced to a well known type: a point, at which the right-hand members of a first-order system are analytic, but vanish simultaneously. The power expansions of these functions about the point start with terms of the third degree. When all further terms are omitted, a system is obtained whose solution is wholly known; unfortunately, no method is at hand for extending its properties to the given system. We finish the chapter with a discussion of the totality of half orbits which reach to infinity, regarded as an *ensemble*.

---

* Presented to the Society, January 2, 1926; received by the editors in April, 1926. In substance this paper is the author's Doctor's thesis, presented at Harvard, 1925.

It is known that for sufficiently large values of the constant of Jacobi there is a closed oval of zero velocity containing the two massive bodies; its exterior is a possible region of motion of the particle. This constitutes the simplest case in which rejection to infinity can occur, and forms the subject of the last two chapters.

Chapter III deals with the *surface of section*, a concept which has become familiar through the writings of Poincaré and Birkhoff. We show that our dynamical problem possesses a ring-shaped surface of section, bounded by a periodic orbit, and by the singularity at infinity. We employ the familiar method of generalizing by continuity from an integrable case. Our results are valid only for sufficiently small values of the ratio of the masses of the heavy bodies.

In Chapter IV, we begin by extending to our case the results concerning the existence and properties of symmetric periodic orbits which were obtained by Birkhoff in the case of motion inside an oval of zero velocity. Then we consider the properties and distribution of the complete orbits which reach to infinity in one or in both directions. As everything here is a consequence of the transformation on the surface of section, it is subject to the same restrictions under which the existence of the latter has been established.

Chapter I. On orbits extending to infinity

1. **The equations of motion.** The restricted problem of three bodies in the plane can, by a proper choice of units of mass, length, and time, be reduced to the study of the motion of a particle P under the Newtonian attraction of two bodies S and J, of masses μ and $1-\mu$, which revolve in circles about their mutual center of gravity O at unit angular velocity. The distances $\overline{OS}$ and $\overline{OJ}$ will then be $1-\mu$ and μ respectively.

Let the fixed XY-axes be chosen through the center of gravity of S and J, in the plane of motion, and in such a manner that S and J revolve in the positive direction (Fig. 1). The equations of motion of $P:(X, Y)$ will be

$$\frac{d^2X}{dt^2} = -\frac{\partial \Pi}{\partial X}, \qquad \frac{d^2Y}{dt^2} = -\frac{\partial \Pi}{\partial Y}, \tag{1}$$

$$\Pi = \Pi(X,Y,t) = -\frac{1-\mu}{r_1} - \frac{\mu}{r_2}.$$

In polar coördinates, $P:(r, \theta)$:

$$\frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2 = -\frac{\partial \Pi}{\partial r}, \qquad \frac{d}{dt}\left(r^2\frac{d\theta}{dt}\right) = -\frac{\partial \Pi}{\partial \theta}. \tag{1'}$$

If the rotating xy-axes are introduced (Fig. 1), with respect to which S and J are fixed and of zero ordinate, the equations may be written in the well known Jacobi form:

$$\frac{d^2x}{dt^2} - 2\frac{dy}{dt} = \frac{\partial\Omega}{\partial x}, \quad \frac{d^2y}{dt^2} + 2\frac{dx}{dt} = \frac{\partial\Omega}{\partial y},$$

$$\Omega = \Omega(x,y) = \frac{1}{2}r^2 - \Pi = \frac{1}{2}r^2 + \frac{1-\mu}{r_1} + \frac{\mu}{r_2}, \tag{2}$$

$$r_1 = [(x-\mu)^2 + y^2]^{1/2}, \quad r^2 = [(x+1-\mu)^2 + y^2]^{1/2},$$

$$r = (x^2 + y^2)^{1/2}.$$

They admit the integral of Jacobi

$$v^2 = \left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 = 2\Omega - C. \tag{3}$$

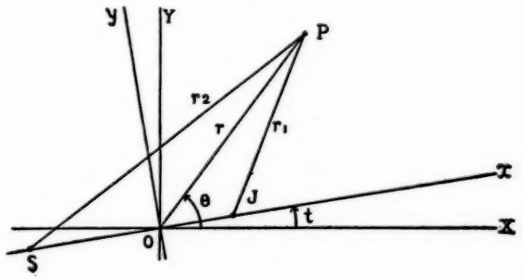

Fig. 1

Our problem does not in general possess an integral of energy; hence we cannot speak of the "energy of the system." It will, however, be useful to introduce the quantity η, which will be called the *specific energy of the particle* P:

$$\eta = \tfrac{1}{2}V^2 + \Pi$$

$$= \frac{1}{2}\left[\left(\frac{dx}{dt} - y\right)^2 + \left(\frac{dy}{dt} + x\right)^2\right] + \Pi, \tag{4}$$

$$V^2 = \left(\frac{dX}{dt}\right)^2 + \left(\frac{dY}{dt}\right)^2.$$

During a given motion, η evidently changes with the time. Its physical interpretation is $\eta = E/\epsilon$, where E is the total energy that P would have if its mass were ϵ and if S and J were fixed at the positions they occupy at the time t in question.

If we introduce the *double areal velocity* G:

$$G = r^2 \frac{d\theta}{dt},$$

the integral of Jacobi may be written in the form

$$2(G - \eta) = C. \tag{3'}$$

2. **Some theorems on stability.*** When $\mu=0$, we have the problem of two bodies. Then η remains constant as P moves along a given orbit, for the system admits the integral of energy. Its value has a fundamental bearing on the stability of the motion, a fact which follows from the familiar theorem of celestial mechanics *The orbit is hyperbolic, parabolic, or elliptic, according as* $\eta>0$, $\eta=0$, or $\eta<0$. (This will be seen, e.g., from the equation $2\eta G^2=e^2-1$ proved on page 313.) The theorems of this section may be regarded as the generalization of this fact to $\mu>0$.

THEOREM 1. *As long as* $\eta\leqq\eta_0<0$, P *will lie within the rotating equipotential oval* $\Pi=\eta_0$, *and hence within a fixed distance of the origin which depends uniquely on the constant* η_0.

For, from (4), $\Pi\leqq\eta\leqq\eta_0$. As r_1 increases from 0 to $+\infty$, Π increases from $-\infty$ to 0. Hence P lies *within* the oval $\Pi=\eta_0$. The last part of the theorem follows from the known form of the equipotentials about a pair of Newtonian particles.

We shall now consider one of the halves Ab into which an orbit is divided by a point A upon it (Fig. 2). The sense in which the particle P describes Ab does not matter. We orient Ab in the sense Ab, and denote by s the arc-length measured from A. Likewise, we take the orientations OX and OP. The angles θ and ψ are measured positively from OX to OP, and from OP to AB, respectively; and $\phi=\theta+\psi$. Let quantities pertaining to A be indicated by subscripts $(_0)$, those pertaining to P, P_n, etc., by similarly marked letters. Finally let $0\leqq\theta_0<2\pi$.

THEOREM 2. *On the hypothesis that*

(a) $r_0\geqq 5$; $(>(5+17^{1/2})/2)$;

(b) *either* $0\leqq\psi_0\leqq\pi/2$, *or else* $3\pi/2\leqq\psi_0<2\pi$;

(c) *the specific energy* η *of the particle* P *of which* Ab *is the orbit is positive or zero at every point of this curve; we conclude that*

(α) Ab *extends to infinity without apsides* (*i.e.*, $\psi\neq\pi/2$ *or* $3\pi/2$);

* Cf. the note at the end of this section.

(β) $\lim_{s\to\infty}\theta$ *exists* $(=\Theta)$;
(γ) $\lim_{s\to\infty}\psi=0$;
(δ) $\lim_{s\to\infty}\eta=\lim_{s\to\infty}V^2/2=\eta_\infty\geqq 0$ *and finite.*

In our proof we shall first examine the analytic nature of the curve Ab and the quantities r, θ, s, ϕ, ψ. This is necessary in order to justify the use of many of the formulas of polar coördinates, etc. Then we shall establish some inequalities, (5), (6), (6′), of great importance in this and later theorems. Finally, we shall prove successively the conclusions (β), (γ), (α), (δ).

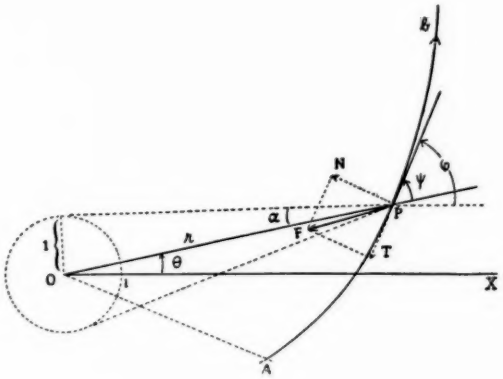

FIG. 2

Let us denote by $\{Ab\}$ the connected portion of Ab which starts at A and remains outside the circle $r=5$. It follows from well known existence theorems applied to the equations (1) that $\{AB\}$ may be represented as

$$X = F_1(t), \qquad Y = F_2(t),$$

where $F_1(t)$ and $F_2(t)$ are analytic and single-valued for t real and $t_1 \leqq t \leqq t_2$ (t_1 may be the symbol $-\infty$, or t_2 the symbol $+\infty$,—the equality sign, naturally, being omitted). Hence $\{Ab\}$ may be represented as

$$r = f_1(t), \qquad \theta = f_2(t),$$

and $f_1(t)$, $f_2(t)$ will be of the same nature as $F_1(t)$, $F_2(t)$. From hyp. (c) and formula (4) it follows that $ds/dt(=V)$ cannot vanish at a finite distance from the origin. Hence, $\{Ab\}$ is an analytic curve without cusps.

The arc, $s=s(t)$, is clearly analytic in t for the values considered. Since $ds/dt \neq 0$, s may be introduced as parameter, when we shall have

$$r = r(s), \quad \theta = \theta(s), \quad \psi = \psi(s), \quad \phi = \theta + \psi = \phi(s),$$

all these functions being single-valued and analytic for all real s in the interval $0 \leqq s < s_1$ (s_1 may be the symbol $+\infty$). The same may be said of the curvature $\kappa = |d\phi/ds|$, except where $\kappa = 0$, when this function is merely continuous.

We now turn to the inequalities. Let F be the absolute value of the resultant PF of the forces acting on P, and N the absolute value of its normal component to the path, T the algebraic value of its tangential component. Since S and J are always within the unit circle of center O, we have

$$F \leqq \frac{1-\mu}{r_1^2} + \frac{\mu}{r_2^2} \leqq \frac{1-\mu}{(r-1)^2} + \frac{\mu}{(r-1)^2} = \frac{1}{(r-1)^2}.$$

Further, as regards the line of action of the resultant, the angle FPN lies between $\pi/2-\psi-\alpha$ and $\pi/2-\psi+\alpha$, where $\alpha = \sin^{-1}(1/r)$ is the angle between the radius-vector and one of the tangents through P to the unit circle. ($\angle FPN$ and α are "small" unoriented angles.) It follows that

$$\begin{aligned} N &= F\,|\cos(\angle FPN)| = F\,|\sin(\psi+\delta)|, \quad |\delta| < \alpha, \\ &= F\,|\sin\psi\cos\delta + \cos\psi\sin\delta|; \end{aligned}$$

therefore

$$\begin{aligned} N &\leqq F(|\sin\psi| \cdot 1 + |\cos\psi| \cdot \sin\alpha) \\ &\leqq F\left(|\sin\psi| + |\cos\psi| \cdot \frac{1}{r}\right). \end{aligned}$$

Hence, from the previous inequality,

$$N \leqq \frac{|\sin\psi|}{(r-1)^2} + \frac{|\cos\psi|}{r(r-1)^2}.$$

The normal resolution of force and acceleration gives, by elementary mechanics,

$$\kappa V^2 = N,$$

therefore

$$\kappa V^2 \leqq \frac{|\sin\psi|}{(r-1)^2} + \frac{|\cos\psi|}{r(r-1)^2}.$$

By hyp. (c),

$$\eta = \frac{V^2}{2} + \Pi \geqq 0,$$

therefore

$$V^2 \geqq -2\Pi.$$

Considerations like those of the preceding paragraph lead to the inequality

$$-2\Pi = 2\left(\frac{1+\mu}{r_1} + \frac{\mu}{r_2}\right) \geqq \frac{2}{r+1},$$

therefore

$$V^2 \geqq \frac{2}{r+1};$$

whence

$$\kappa \leqq \frac{r+1}{2(r-1)^2}|\sin\psi| + \frac{r+1}{2r(r-1)^2}|\cos\psi|, \tag{5}$$

which is the first of the inequalities we are seeking.

A first result of this inequality is that the curvature of $\{Ab\}$ at an apsis (r_1, θ_1) is less than the curvature of the circle of radius r_1 (not r_1 of Fig. 1). For, by (5), when $\psi = \pi/2$,

$$\kappa_1 \leqq \frac{r_1+1}{2(r_1-1)^2} = \frac{1}{r_1}\,\frac{r_1(r_1+1)}{2(r_1-1)^2},$$

and a little computation shows that when $r \geqq 5$,

$$\frac{r_1(r_1+1)}{2(r_1-1)^2} < 1,$$

therefore

$$\kappa_1 < \frac{1}{r_1}.$$

It follows that $r(s)$ can have no maxima, i.e., $\{Ab\}$ *has no higher apsides.* This fact, which is almost geometrically evident from the last inequality, is proved by considering the osculating circle C drawn at a point (r_1, θ_1) where we suppose $dr/ds = 0$. It is tangent externally to the circle K of radius r_1, center at O. Since $\{Ab\}$ has contact of higher order with C, it will lie outside K in the neighborhood of (r_1, θ_1). Hence (r_1, θ_1) is not a point of maximum r. It follows from this, and from hyp. (a), (b), that $r(s)$ increases monotonically with s. Hence $\{Ab\}$ is identical with Ab, and all our above results, obtained for the former curve, apply to the whole half orbit.

The length of Ab is infinite. For if Ab were within a fixed distance of the origin, there would exist a constant, $h>0$, such that $|ds/dt|>h$ (by (4) and hyp. (c)). Hence the length of Ab would be infinite. And of course it would be infinite if the contrary assumption were made about Ab.

We return to the inequality (5). Recalling the elementary formulas of polar coördinates:

$$\sin\psi = r\frac{d\theta}{ds}, \quad \cos\psi = \frac{dr}{ds},$$

which are valid through Ab, by the nature of this curve, we have

$$\left|\frac{d\phi}{ds}\right| = \kappa \leqq \frac{r+1}{2(r-1)^2}|\sin\psi| + \frac{r+1}{2r(r-1)^2}|\cos\psi|,$$

therefore

$$|d\phi| \leqq r\frac{(r+1)}{2(r-1)^2}|d\theta| + \frac{r+1}{2r(r-1)^2}|dr|.$$

We shall take, as we may, $ds>0$. Because of the monotonic increase of r, $dr\geqq 0$, and $|\psi|\leqq\pi/2$. Consider first the case where $d\theta\geqq 0$ between two points (r_1, θ_1) and (r_2, θ_2), $r_1\leqq r_2$. We are lead to

$$d\phi \leqq |d\phi| < \frac{r(r+1)}{2(r-1)^2}d\theta + \frac{r+1}{2r(r-1)^2}dr;$$

$$d\psi = d\phi - d\theta < \left[\frac{r(r+1)}{2(r-1)^2} - 1\right]d\theta + \frac{r+1}{2r(r-1)^2}dr.$$

A little computation shows that the bracket decreases as r increases when $r\geqq 5$, so that

$$\frac{r(r+1)}{2(r-1)^2} - 1 \leqq -\frac{1}{16}, \quad \text{when } r \geqq 5.$$

Hence

$$d\psi \leqq -\frac{1}{16}d\theta + \frac{r+1}{2r(r-1)^2}dr.$$

Taking the definite integral of this along Ab from (r_1, θ_1) to (r_2, θ_2),

$$\psi_2 - \psi_1 \leqq -\frac{1}{16}(\theta_2 - \theta_1) + \left\{\frac{1}{r_1 - 1} - \frac{1}{r_2 - 1} + \frac{1}{2}\log\left(\frac{r_2}{r_1}\right)\frac{r_1 - 1}{r_2 - 1}\right\};$$

since $r_2 \geqq r_1 \geqq 5$, this yields

$$\psi_2 - \psi_1 \leqq -\frac{1}{16}(\theta_2 - \theta_1) + \frac{1}{r_1 - 1} - \frac{1}{r_2 - 1}. \tag{6}$$

Turn now to the case where $d\theta \leqq 0$ between (r_1, θ_1) and (r_2, θ_2). In the work up to now we have made no use of the fact that S and J are revolving about O in the "counterclockwise" direction. There is, then, nothing to distinguish the formulas we have obtained from those we should have found had we measured all angles positively in the "clockwise" sense. Accordingly, our last result (6) must hold when θ, ϕ, ψ are replaced by their negatives, and the condition $d\theta \geqq 0$ by $d(-\theta) \geqq 0$, i.e., $d\theta \leqq 0$. Hence, after multiplying through by -1, we get

$$\psi_2 - \psi_1 \geqq -\frac{1}{16}(\theta_2 - \theta_1) - \frac{1}{r_1 - 1} + \frac{1}{r_2 - 1}. \tag{6'}$$

This formula replaces (6) when θ decreases monotonically between (r_1, θ_1) and (r_2, θ_2).

We are now in a position to prove the conclusion (β). Assume first that from a certain point (r_1, θ_1) on, θ increases monotonically with s. From (6) we have, since $r_1 \geqq 5$,

$$\psi_2 - \psi_1 \leqq -\frac{1}{16}(\theta_2 - \theta_1) + \frac{1}{4}.$$

Now $0 \leqq \psi_1 \leqq \pi/2$, $0 \leqq \psi_2 \leqq \pi/2$, and hence

$$-\frac{\pi}{2} \leqq -\frac{1}{16}(\theta_2 - \theta_1) + \frac{1}{4};$$

therefore

$$\theta_2 \leqq \theta_1 + 4 + 8\pi.$$

Thus, θ_2 (or θ) increases steadily but remains finite as $s_2 \to \infty$. Hence it approaches a limit, Θ, as we wished to prove. The same conclusion is arrived at when θ decreases monotonically.

The only case where θ does not behave as we have assumed is when it passes through infinitely many maxima and minima as $s \to \infty$. Let these occur at P_{2k} and P_{2k-1} respectively ($k = 1, 2, \cdots$). At these points the angle ψ will be zero. Hence, by (6), (6'),

$$|\theta_{2k} - \theta_{2k-1}| \leqq 16\left(\frac{1}{r_{2k-1} - 1} - \frac{1}{r_{2k} - 1}\right),$$

$$|\theta_{2k+1} - \theta_{2k}| \leqq 16\left(\frac{1}{r_{2k} - 1} - \frac{1}{r_{2k+1} - 1}\right).$$

Let P' and P'' be an arbitrary pair of points on Ab, $s'<s''$. If $\theta'<\theta''$, call P_{2p-1} the minimum adjacent to (or coincident with) P', and P_{2q} the maximum adjacent to (or coincident with) P''. If $\theta'>\theta''$, call P_{2p} the maximum adjacent to P', and P_{2q-1} the minimum adjacent to P''. It will be seen that as s' and s'' become infinite, the s for the other corresponding points do likewise. Also, $p \leqq q$. In the first case, $(\theta'<\theta'')$, we have

$$|\theta''-\theta'| \leqq |\theta_{2q}-\theta_{2p-1}| \leqq |\theta_{2p}-\theta_{2p-1}|+|\theta_{2p+1}-\theta_{2p}|+\cdots+|\theta_{2q}-\theta_{2q-1}|$$

$$\leqq 16\left\{\frac{1}{r_{2p-1}-1}-\frac{1}{r_{2p}-1}+\frac{1}{r_{2p}-1}-\frac{1}{r_{2p+1}-1}+\cdots-\frac{1}{r_{2q}-1}\right\},$$

therefore

$$|\theta''-\theta'| \leqq 16\left\{\frac{1}{r_{2p-1}-1}-\frac{1}{r_{2q}-1}\right\}.$$

Similarly, in the case $\theta'>\theta''$,

$$|\theta''-\theta'| \leqq 16\left\{\frac{1}{r_{2p}-1}-\frac{1}{r_{2q-1}-1}\right\}.$$

We have seen that r increases monotonically with s. Hence, as $s\to\infty$, either $r\to\infty$, or else $\lim r=R\geqq 5$. In either case the braces in the last two inequalities approach zero. Thus

$$\lim |\theta''-\theta'| = 0$$

as s' and s'' become infinite independently of one another. It follows from a fundamental theorem of analysis that

$$\lim_{s\to\infty}\theta = \Theta,$$

and thus conclusion (β) is established in all cases.

Turn now to the conclusion (γ) or our theorem. When θ changes monotonically with s from a certain point on, it follows from (6) or (6′) that $\lim|\psi''-\psi'|=0$ as s' and s'' both become infinite; hence, $\lim_{s\to\infty}\psi$ exists. *Assume it distinct from zero.* By the formula $rd\theta/dr=\tan\psi$ it follows that

$$\left|r\frac{d\theta}{dr}\right| > a > 0$$

from a certain point on; therefore

$$|\theta| > a\log r + b \qquad (a,b \text{ const.}).$$

If $r \to \infty$ with s, we are led to a contradition with (β). If r does not behave in this way, $\lim_{s\to\infty} r = R-$, and the formula $r d\theta/ds = \sin\psi$ leads to

$$\left|\frac{d\theta}{ds}\right| = \frac{|\sin\psi|}{R} > A > 0$$

from a certain point on; therefore

$$|\theta| > As + B \qquad (A, B \text{ const.}).$$

Again we are brought into conflict with (β). Thus, (γ) is proved under the assumption of the monotonic change of θ.

If θ is not monotonic, it will have an infinite succession of maxima and minima, at the points P_{2k} and P_{2k-1} respectively. Between P_{2k-1} and P_{2k}, $\psi \geqq 0$; let $\psi^{(2k)}$ be its maximum maximorum there, attained at the point $P^{(2k)}$: $s_{2k-1} \leqq s^{(2k)} \leqq s_{2k}$. By (6),

$$(0 \leqq)\ \psi^{(2k)} \leqq -\frac{1}{16}(\theta^{(2k)} - \theta_{2k-1}) + \frac{1}{r_{2k-1} - 1} - \frac{1}{r^{(2k)} - 1}.$$

Similarly, for the minimum minimorum $\psi^{(2k+1)}$ between P_{2k} and P_{2k+1}, we have

$$(0 \geqq)\ \psi^{(2k+1)} \geqq -\frac{1}{16}(\theta^{(2k+1)} - \theta_{2k}) - \frac{1}{r_{2k} - 1} + \frac{1}{r^{(2k+1)} - 1}.$$

Hence it is readily seen that

$$\lim_{k\to\infty} \psi^{(2k)} = \lim_{k\to\infty} \psi^{(2k+1)} = 0,$$

and the truth of the conclusion (γ) is evident.

The proof of (α) now follows from the formula $dr/ds = \cos\psi$. For we see from (γ) that

$$\frac{dr}{ds} > M > 0$$

from a certain point on; therefore

$$r > Ms + N \qquad (M, N \text{ const.}),$$

which makes (α) evident.

Conclusion (δ) follows from the formula of elementary mechanics giving the tangential resolution of force and acceleration of P (Fig. 2):

$$\frac{dV}{dt} = \frac{d}{ds}\left(\frac{V^2}{2}\right) = T.$$

It is clear from what we have proved regarding the geometric nature of Ab that, from a certain point on, $T<0$. Hence V^2 decreases monotonically as $s\to\infty$. This renders the truth of (δ) obvious, and thus the proof of Theorem 2 is complete.

It is interesting to state these results in terms of the *osculating eccentricity*, e. Theorem 1 shows that as long as $e\leqq 1-h$, h an arbitrarily small positive constant, the orbit remains within a fixed distance of the origin,* while in Theorem 2, we may replace the hyp. (c) by the requirement that $e\geqq 1$ along Ab. We do not go into further detail, because there is a slight variance in usage in the definition of e. Once its meaning is fixed upon, the connection may be made with the value of η by means of formulas analogous to the one cited in the opening paragraph of this section.

Theorem 2 cannot be applied directly to the testing of the stability of a half orbit given by its initial conditions, because it requires a knowledge concerning the value of η throughout the entire length. But it is easy to derive a test which is directly applicable:

THEOREM 3. *The conclusion of Theorem 2 subsists when the hypothesis* (c) *is replaced by the requirement that, at* A,

$$(c') \qquad \left(\frac{dr}{dt}\right)_0^2 \geqq \frac{2}{r_0-1}.$$

From hyp. (c) it follows that a segment AB of Ab exists within which $|dr/dt|>0$. Letting (r, θ) denote any point on AB, we clearly have

$$\frac{d^2r}{dt^2}=\frac{dr}{dt}\frac{d}{dr}\left(\frac{dr}{dt}\right)=\frac{1}{2}\frac{d}{dr}\left(\frac{dr}{dt}\right)^2.$$

By (1′), and Fig. 2,

$$\frac{d^2r}{dt^2}=r\left(\frac{d\theta}{dt}\right)^2-\frac{\partial\Pi}{\partial r}\geqq -\,[\text{greatest possible } F]\geqq\frac{-1}{(r-1)^2};$$

therefore

$$\frac{d}{dr}\left(\frac{dr}{dt}\right)^2\geqq\frac{-2}{(r-1)^2}.$$

Integrating from $A:(r_0, \theta_0)$ to $B:(r_1, \theta_1)$, with r as independent variable, $dr>0$, we obtain

* This is related to the classical theorem of celestial mechanics about the *stability of the major axis.*

$$\left(\frac{dr}{dt}\right)_1^2 - \frac{2}{r_1 - 1} \geqq \left(\frac{dr}{dt}\right)_0^2 - \frac{2}{r_0 - 1} \geqq 0,$$

the second inequality coming from hyp. (c′).

It follows that throughout Ab, $|dr/dt| > 0$. Otherwise, on letting B approach the first point where $dr/dt = 0$, our preceding formula would lead to a discontinuity in the function dr/dt, contrary to the nature which the functions r, θ, etc. have been seen to possess on Ab (p. 292). Hence, for all Ab,

$$2\eta = \left(\frac{dr}{dt}\right)^2 + r^2\left(\frac{d\theta}{dt}\right)^2 + 2\Pi \geqq \left(\frac{dr}{dt}\right)^2 - \frac{2}{r - 1},$$

and from the preceding inequality,—which is true throughout Ab—the right-hand member is positive or zero. Thus, $\eta \geqq 0$ throughout Ab, and all the hypotheses of Theorem 2 are satisfied, which proves our theorem.

THEOREM 4. *Let h be any positive constant. Then there exist half orbits extending to infinity with $\eta_\infty > h$, and others with $\eta_\infty < h$.*

We can always take $(dr/dt)_0^2$ so large that

$$\left(\frac{dr}{dt}\right)_0^2 - \frac{2}{r_0 - 1} > 2h,$$

the other hypotheses of Theorem 3 being realized. Then $\eta_\infty > h$. To prove the second part of the theorem, take

$$\left(\frac{dr}{dt}\right)_0^2 = \frac{2}{r_0 - 1},$$

and suppose the other hypotheses of Theorem 3 fulfilled. We may further suppose that r_0 is so large, and $(d\theta/dt)_0$ so small, that

$$\eta_0 + \left(\frac{2}{r_0 - 1}\right)^{1/2} < h.$$

This is clearly possible, since the left-hand member is equal to

$$\frac{1}{r_0 - 1} + \frac{1}{2}r_0^2\left(\frac{d\theta}{dt}\right)_0^2 + \Pi_0 + \left(\frac{2}{r_0 - 1}\right)^{1/2}.$$

Then, I say, this orbit (which extends to infinity, by Theorem 3) will have $\eta_\infty < h$. For, as in the proof of Theorem 3,

$$\left(\frac{dr}{dt}\right)^2 - \frac{2}{r - 1} \geqq 0.$$

If $dr/dt>0$,

$$dt \leqq \left(\frac{r-1}{2}\right)^{1/2} dr.$$

By (1′), the expression $G=r^2 d\theta/dt$, and (3′),

$$d\eta = dG \leqq \frac{dt}{(r-1)^2} \leqq \frac{1}{2^{1/2}}(r-1)^{-3/2}dr.$$

Integrating from P_0 to infinity along Ab,

$$\eta_\infty \leqq \eta_0 + \left(\frac{2}{r_0-1}\right)^{1/2} < h.$$

Similarly for $dr/dt<0$.

COROLLARY. *In Theorem* 4, *the value of the constant of Jacobi for the half orbits may be chosen arbitrarily.*

For it will be seen that we can satisfy our various hypotheses for any given value of C.

Note. All the work of this section applies almost verbatim to the motion of a particle P under the Newtonian attraction of arbitrarily distributed masses, animated with an arbitrary continuous motion, provided merely that their total mass be unity, and that they remain within unit distance of the origin. The only alterations are that $\{Ab\}$ and Ab will not necessarily be analytic, but will have a continuously varying curvature. Naturally, everything about the integral of Jacobi falls out.

3. **The character of the orbits in the neighborhood of infinity.** We shall consider the half orbit Ab, with the same notation as before (Fig. 2). The leading theorem here is the following:

THEOREM 5. *If Ab extends to infinity, it does so with its direction from the origin, the direction of its tangent, and its specific energy all approaching definite limits. In other words, if* $\lim_{s\to\infty} r = \infty$, *the remaining conclusions of Theorem* 2 *are in force.*

If at a single point A' of Ab the hypotheses of Theorem 3 were realized for the curve $A'b$, the conclusion of our theorem would follow at once from the latter. We shall assume from now on that such a point A' does not occur. Then, whenever $r(s)$ increases with s, we shall need to have (after $r>5$)

$$\left(\frac{dr}{dt}\right)^2 < \frac{2}{r-1}.$$

When $r(s)$ decreases (s increasing) from a maximum at P_1, this inequality must still be true; otherwise Theorem 3 would deny the existence of P_1. Thus this inequality will be valid throughout Ab. Therefore

$$\lim_{s\to\infty} \frac{dr}{dt} = 0.$$

On substituting the polar expression for η in (3′), we have

$$2G - \left(\frac{dr}{dt}\right)^2 - r^2\left(\frac{d\theta}{dt}\right)^2 - 2\Pi = C.$$

Eliminating $d\theta/dt$ by means of the definition of G,

$$G = r^2 \frac{d\theta}{dt},$$

we obtain the quadratic equation in G:

$$G^2 - 2r^2G + r^2\left[2\Pi + C + \left(\frac{dr}{dt}\right)^2\right] = 0.$$

This determines the two values of G, G' and G'':

$$(7)\qquad \begin{aligned} G' &= r^2 + \left(r^4 - r^2\left[2\Pi + C + \left(\frac{dr}{dt}\right)^2\right]\right)^{1/2}, \\ G'' &= r^2 - \left(r^4 - r^2\left[2\Pi + C + \left(\frac{dr}{dt}\right)^2\right]\right)^{1/2}. \end{aligned}$$

From the equation proved in the paragraph before last, it follows that from a certain point of Ab on, $G=G''$. Otherwise, by (7), $G\to\infty$ as $s\to\infty$. Hence, by (3′), $\eta\to\infty$, and we should be able to apply Theorem 2,—which would give the contradictory result that η approaches a limit.

We wish to evaluate $\lim_{s\to\infty} G$, if it exists. Assume, as we may, that r is large, and expand the expression for G'' in (7) by means of the binomial theorem. A little computation shows that

$$\lim_{s\to\infty} G = \frac{C}{2};$$

hence, by (3′),

$$\lim_{s\to\infty} \eta = \lim_{s\to\infty} \frac{V^2}{2} = 0.$$

Now reason with the first equation of (1′) as we did in the proof of Theorem 3, only with the inequality signs reversed. Here we have, since

$$G = r^2\frac{d\theta}{dt}, \qquad \lim_{s\to\infty} G = \frac{C}{2},$$

therefore

$$\frac{d^2r}{dt^2} \leqq \frac{G^2}{r^3} - \frac{1}{(r+1)^2}$$

$$\leqq \frac{a}{r^3} - \frac{1}{(r+1)^2} \qquad (a, \text{ const.}).$$

Now, by Fig. 2,

$$-\frac{\partial \Pi}{\partial r} \leqq -F\cos\alpha \leqq -\frac{1}{(r+1)^2}\left(1-\frac{1}{r^2}\right)^{1/2};$$

therefore

$$\frac{d^2r}{dt^2} \leqq \frac{a}{r^3} - \frac{1}{(r+1)^2}\left(1-\frac{1}{r^2}\right)^{1/2}.$$

Since by hypothesis Ab goes to infinity, from a certain point on, the right-hand member of this inequality will be negative. But $d^2r/dt^2<0$ implies that r is a *maximum* at every apsis. It follows that, after a certain point, $|dr/dt|>0$. The preceding inequality shows, further, that there will be a positive constant K such that, eventually,

$$\frac{1}{2}\frac{d}{dr}\left(\frac{dr}{dt}\right)^2 = \frac{d^2r}{dt^2} < -\frac{K}{(r+1)^2}.$$

Integrating along the orbit, from (r, θ) to infinity, we get

$$V^2 \geqq \left(\frac{dr}{dt}\right)^2 \geqq \frac{2K}{r+1}, \qquad K>0.$$

From this we may pass to the equations (6), (6′), just as we did in the proof of Theorem 2 from the inequality

$$V^2 \geqq \frac{2}{r+1},$$

the only difference being that the result is valid only at a somewhat greater distance from the origin. Thus, all the material is at hand for the proof of the conclusion of Theorem 2, and our theorem is established.

We turn now to some theorems concerning the existence of an asymptote for a half orbit Ab reaching to infinity. It has been proved that when $s\to\infty$, the limits of V, G, and ϕ exist. Denote them respectively by V_∞, G_∞, and Φ. Let homogeneous coördinates (x_1, x_2, x_3) be introduced in the XY-plane:

$$X = \frac{x_1}{x_3}, \qquad Y = \frac{x_2}{x_3}.$$

The distance from the origin to the tangent to Ab at any point (r, θ) is clearly

$$|\, r \sin \psi \,| = \left| r^2 \frac{d\theta}{ds} \right| = \left| \frac{G}{V} \right|.$$

This tangent makes an angle ϕ with the X-axis. The use of Hesse's normal form, and an inspection of signs in the various cases, give us the equation of the tangent in the running coördinates (x_1, x_2, x_3):

$$V \sin \phi \cdot x_1 - V \cos \phi \cdot x_2 - G \cdot x_3 = 0.$$

When $s \to \infty$, the coefficients approach those of the equation

$$V_\infty \sin \Phi \cdot x_1 - V_\infty \cos \Phi \cdot x_2 - G_\infty \cdot x_3 = 0. \tag{8}$$

This equation becomes illusory when $V_\infty = G_\infty = 0$. But in that case, by (3′), $C = 0$, and

$$\frac{G}{V} = \frac{V}{2} + \frac{\Pi}{V}.$$

From the next to the last inequality in the proof of Theorem 4, and other previous results, we have

$$V|\,| \geqq \left(\frac{2K}{r+1} \right)^{1/2}, \text{ and } |\,\Pi\,| \leqq \frac{1}{r-1};$$

therefore

$$\lim_{s \to \infty} \frac{G}{V} = 0.$$

Letting G_∞ / V_∞ stand for this limit, we see that (8) can always be given a meaning. With this interpretation we have

THEOREM 6. *The half orbit Ab approaches the line* (8) *as asymptote.*

We have merely to show that (8) is tangent to Ab at a point on the line infinity. There is no loss of generality in assuming that $\Phi = 0$. By means of the non-singular projective transformation of the XY-plane into the $X'Y'$-plane

$$x_1' = x_3, \quad x_2' = x_2, \quad x_3' = x_1, \tag{$\mathfrak{T}$}$$

we carry the line infinity into the Y' axis, and (8) into

$$V_\infty \sin \Phi \cdot x_3' - V_\infty \cos \Phi \cdot x_2' - G_\infty \cdot x_1' = 0, \tag{8′}$$

while Ab goes into what we shall call $A'b'$. We wish to show that (8′) is tangent to $A'b'$ at a point on the Y'-axis. The equation of the tangent to $A'b'$ at an arbitrary point is obtained by applying ($\mathfrak{T}$) to the equation of the tangent to Ab (since projection conserves tangency). Thus we get

$$V \sin \phi \cdot x_3' - V \cos \phi \cdot x_2' - G \cdot x_1' = 0.$$

As $s \to \infty$, its coefficients approach those of (8′). We have merely to interpret this statement in non-homogeneous coördinates to see that this means that (8′) is tangent to $A'b'$ at a point on the X'-axis (the origin, as we see from Theorem 4), and our theorem is proved.

COROLLARY. *If $G_\infty \neq 0$, Ab approaches an asymptote not passing through the origin when $V_\infty \neq 0$, while it is tangent to the line infinity when $V_\infty = 0$. If $G_\infty = 0$, Ab always has an asymptote passing through the origin.*

All these theorems show how greatly the motion along an orbit reaching to infinity resembles the hyperbolic and parabolic motion in the problem of two bodies. The new feature in the present case is the conceivability of the particle's oscillating about its asymptote. This might be expected, in view of the oscillatory nature of the field of force.

Note. Here again we observe that our work applies verbatim to a much more general class of cases than the restricted problem of three bodies. It applies to those systems mentioned in the note at the close of the last section which possess an integral of Jacobi (3) (where $-\Pi$ is the force function).

CHAPTER II. ANALYTIC STUDY OF INFINITY

1. **A transformation of the equations of motion.** We shall consider the equations of motion in the rotating xy-plane. When the variables $x' = dx/dt$, $y' = dy/dt$ are introduced, the equations I, (2), reduce to a system of the first order:

$$(1) \qquad \begin{aligned} \frac{dx}{dt} &= x', \\ \frac{dy}{dt} &= y', \\ \frac{dx'}{dt} &= 2y' + x - \frac{\partial \Pi}{\partial x}, \\ \frac{dy'}{dt} &= -2x' + y - \frac{\partial \Pi}{\partial y}. \end{aligned}$$

These define in general a unique integral curve through each point of the $xyx'y'$-space. The same is clearly true of the 3-dimensional manifold composed of the totality of points (x, y, x', y') corresponding to a given value of the constant of Jacobi, C, and determined by equation (3) of Chapter I. In either case, points for which (x, y) is at infinity form a singularity of the integral curves, as we see on examination of (1). It is our object in this section to reduce the singularity to a standard type by means of a change of variables.

We shall perform the transformation

$$x + (-1)^{1/2}y = \frac{1}{(\alpha + (-1)^{1/2}\beta)^2}, \tag{2}$$

or

$$x = \frac{\alpha^2 - \beta^2}{(\alpha^2 + \beta^2)^2}, \quad y = -\frac{2\alpha\beta}{(\alpha^2 + \beta^2)^2}, \quad x^2 + y^2 = \frac{1}{(\alpha^2 + \beta^2)^2}.$$

After a rather lengthy but straightforward computation, which can be simplified somewhat by the use of the rudiments of the theory of complex variables, equations (2) of Chapter I yield

$$\frac{d^2\alpha}{dt^2} = -\frac{1}{2}\left(\alpha - 4\frac{d\beta}{dt}\right) + \frac{3}{\alpha^2 + \beta^2}\left\{\alpha\left[\left(\frac{d\alpha}{dt}\right)^2 - \left(\frac{d\beta}{dt}\right)^2\right] + 2\beta\frac{d\alpha}{dt}\frac{d\beta}{dt}\right\} + \frac{1}{4}(\alpha^2 + \beta^2)^3\Pi_\alpha,$$

$$\frac{d^2\beta}{dt^2} = -\frac{1}{2}\left(\beta + 4\frac{d\alpha}{dt}\right) + \frac{3}{\alpha^2 + \beta^2}\left\{-\beta\left[\left(\frac{d\alpha}{dt}\right)^2 - \left(\frac{d\beta}{dt}\right)^2\right] + 2\alpha\frac{d\alpha}{dt}\frac{d\beta}{dt}\right\} + \frac{1}{4}(\alpha^2 + \beta^2)^3\Pi_\beta,$$

while the integral of Jacobi, I, (3), becomes

$$\alpha^2 + \beta^2 - 4\left[\left(\frac{d\alpha}{dt}\right)^2 + \left(\frac{d\beta}{dt}\right)^2\right] = (\alpha^2 + \beta^2)^3(2\Pi + C).$$

On introducing the new variables α', β', τ, defined by

$$\alpha' = \frac{d\alpha}{dt}, \quad \beta' = \frac{d\beta}{dt}, \quad (\alpha^2 + \beta^2)d\tau = dt,$$

the equations may be written in the form

$$\frac{d\alpha}{d\tau} = (\alpha^2 + \beta^2)\alpha',$$

$$\frac{d\beta}{d\tau} = (\alpha^2 + \beta^2)\beta',$$

$$(3)\quad \frac{d\alpha'}{d\tau} = -\frac{1}{2}(\alpha^2 + \beta^2)(\alpha - 4\beta') + 3\{\alpha(\alpha'^2 - \beta'^2) + 2\beta\alpha'\beta'\} + \frac{1}{4}(\alpha^2+\beta^2)^4\frac{\partial \Pi}{\partial \alpha},$$

$$\frac{d\beta'}{d\tau} = -\frac{1}{2}(\alpha^2+\beta^2)(\beta+4\alpha') + 3\{-\beta(\alpha'^2 - \beta'^2) + 2\alpha\alpha'\beta'\} + \frac{1}{4}(\alpha^2+\beta^2)^4\frac{\partial \Pi}{\partial \beta},$$

with the integral of Jacobi

$$(4)\qquad \alpha^2 + \beta^2 - 4(\alpha'^2 + \beta'^2) = (\alpha^2 + \beta^2)^3(2\Pi + C).$$

The specific energy η is likewise found to be

$$(5)\qquad \eta = \frac{(2\alpha' + \beta)^2 + (2\beta' - \alpha)^2}{2(\alpha^2 + \beta^2)^3} + \Pi.$$

It is seen from (2) that the figure in the $xyx'y'$-space corresponding to the singularity at infinity has gone into the plane $\alpha=\beta=0$ in the $\alpha\beta\alpha'\beta'$-space. The correspondence between points (x, y, x', y') and $(\alpha, \beta, \alpha', \beta')$ is one-to-two, and analytic except when (α, β) is at the origin or infinity. Furthermore, the right-hand members of (3) vanish at $\alpha=\beta=0$. Finally, they are analytic in the neighborhood of $(0, 0, 0, 0)$; for

$$-\Pi = \frac{1-\mu}{r_1} + \frac{\mu}{r_2},$$

and by (2)

$$r_1^{-1} = [(x-\mu)^2 + y^2]^{-1/2} = (\alpha^2+\beta^2)[1 - 2\mu(\alpha^2 - \beta^2) + \mu^2(\alpha^2+\beta^2)^2]^{-1/2},$$

which is clearly analytic at $\alpha=\beta=0$. Likewise for r_2^{-1}. Thus $\partial\Pi/\partial\alpha$ and $\partial\Pi/\partial\beta$ are analytic at $\alpha=\beta=0$, and their power series expansions about that point start with terms of the first degree. This proves our statement.

Thus the singularity we are considering has been reduced to a well known type: a point, at which the right-hand members of the system of differential equations are analytic, but vanish simultaneously. The singularity is isolated in the manifold (4), for the only point of that manifold where $\alpha=\beta=0$ is $(0, 0, 0, 0)$.

The expansions of the right-hand members of (3) at the origin start with terms of the third degree. If the higher terms are removed, we obtain a

system whose solution is completely known; for it is obtained by setting $\Pi \equiv 0$, and hence corresponds to the motion of the particle P under the action of no forces. In particular, it will admit the integral of energy (5) ($\Pi = 0$).

It would be of great importance to us to know whether the equations (1) or (3) with Π left in will continue to admit a generalization of this integral. Unfortunately, the standard methods for treating such singularities known at the present time do not apply to this case.

It is possible to find a continuous single-valued function which is a generalization of the integral of energy in a region abutting on our singularity, and which is an integral in the sense that it remains constant along the integral curves. Since this investigation leads us to some further information regarding the unstable orbits, we shall undertake it in the following section.

2. **The limiting energy as uniform integral.** We shall work in the space of the variables (x, y, x', y'), in which the integral curves are determined by the system of equations (1).

Let $\mathfrak{M}$ denote the point set in the real $xyx'y'$-space, each point of which is characterized by the fact that the unique integral curve of (1) which passes through it* goes to infinity when $t \to +\infty$. Let $\overline{\mathfrak{M}}$ be the corresponding point set for $t \to -\infty$. As everything we shall say for $\mathfrak{M}$ applies, mutatis mutandis, to $\overline{\mathfrak{M}}$, we shall confine ourselves in the sequel to the consideration of $\mathfrak{M}$ and the case $t \to +\infty$.

Let $\eta_\infty = \lim_{t \to \infty} \eta$. Then η_∞ is a single-valued function of (x, y, x', y') on $\mathfrak{M}$; it is obtained by finding the limit of η as $t \to +\infty$ along the integral curve through the point (x, y, x', y'). That η_∞ exists, and is positive or zero, is a result of the theorems of Chapter I. We shall denote by $\mathfrak{M}'$ the subset of $\mathfrak{M}$ for which $\eta_\infty > 0$.

Let us inquire into the nature of $\mathfrak{M}'$. First, it will be seen that *if P is a point of $\mathfrak{M}'$, there exists a real neighborhood in the $xyx'y'$-space about P every point of which belongs to $\mathfrak{M}'$.* From the formula (4) of the preceding chapter, and $G = r^2 d\theta/dt$,

$$\left(\frac{dr}{dt}\right)^2 - \frac{2}{r-1} = -\frac{2}{r-1} - \frac{G^2}{r^2} - 2\Pi + 2\eta\,;$$

* It is known that the only points through which the integral curve is not unique have S or J (Fig. 1) as their projection in the xy-plane. But through such a point there is just one integral curve with a given direction when projected upon this plane, and given value of C. So our definition of $\mathfrak{M}$ may be made to go through if we associate a point with a direction in the xy-plane, and a value of C, whenever the projection of the point in this plane coincides with S or J. A similar result is obtained by regularization. In this discussion we assume that the term "integral curve" includes possible equilibrium points.

since $\lim \eta > 0$, it is seen that at a certain point P_0 of the integral curve through P,

$$\left(\frac{dr}{dt}\right)_0^2 - \frac{2}{r_0 - 1} > 0, \quad \left(\frac{dr}{dt}\right)_0 > 0, \quad r_0 > 5.$$

By continuity, there exists a neighborhood $\mathfrak{N}_0$ about P_0 for every point of which

$$\left(\frac{dr}{dt}\right)^2 - \frac{2}{r - 1} > 0, \quad \frac{dr}{dt} > 0, \quad r > 5. \tag{6}$$

Hence, by Chapter I, Theorem 3, $\mathfrak{N}_0$ belongs to $\mathfrak{M}$. But on account of the fact that

$$2\eta \geqq \left(\frac{dr}{dt}\right)^2 - \frac{2}{r - 1},$$

where the right-hand member is positive in $\mathfrak{N}_0$ and *never decreases* as $t \to +\infty$ along a given integral curve (cf. proof of Chapter I, Theorem 3), it follows that $\mathfrak{N}_0$ belongs to $\mathfrak{M}'$. If $\mathfrak{N}$ be the neighborhood about P into which $\mathfrak{N}_0$ is carried by that deformation of the $xyx'y'$-space defined by (1) which carries P_0 into P, it is clear that $\mathfrak{N}$ belongs to $\mathfrak{M}'$. We leave it to the reader to consider the cases where this deformation may be of an exceptional character.

Further, $\mathfrak{M}'$ *is a connected point set.* Let P_1 and P_2 be two points of $\mathfrak{M}'$. The integral curves through these points lie in $\mathfrak{M}'$. They will, as was seen above, eventually both enter a region in which (6) is satisfied. This latter belongs to $\mathfrak{M}'$, and inspection of the formulas reveals that it is connected. Hence, P_1 and P_2 may be connected by a curve of points of $\mathfrak{M}'$. In conclusion we have

THEOREM 1. *The point set $\mathfrak{M}'$ is a 4-dimensional continuum.*

Further information about the boundary points of $\mathfrak{M}$, $\mathfrak{M}'$, $\overline{\mathfrak{M}}$, and $\overline{\mathfrak{M}}'$ will be obtained in §2 of Chapter IV, in those cases where the problem admits a *surface of section.*

THEOREM 2. *The function η_∞ is continuous on $\mathfrak{M}'$.*

From the continuous nature of the deformation defined by (1) in the $xyx'y'$-space, it will be sufficient to establish the theorem at a point, P, in the neighborhood of which the inequalities (6) are verified. For the integral curve through an arbitrary point of $\mathfrak{M}'$ eventually enters such a neighborhood.

Consider the hyperplane perpendicular to the integral curve at P. There is a neighborhood, $\mathfrak{B}$, of P in this hyperplane throughout which (6) is satisfied, and within which the angle made by the integral curves with the hyperplane is greater than a certain positive constant. Let $\mathfrak{F}$ be the tube-like 4-dimensional region composed of the halves of those integral curves through $\mathfrak{B}$ along which $t \to +\infty$. The region $\mathfrak{F}$ clearly reaches to infinity without reëntering its self.

Determine t in $\mathfrak{F}$ so that it is zero upon $\mathfrak{B}$. Then, within this tube $\mathfrak{F}$, $t = \omega(x, y, x', y')$, where ω is a single-valued analytic function.

Let the subscript $(_0)$ indicate values upon $\mathfrak{B}$. Clearly we have

$$\eta_\infty - \eta_0 = \int_0^\infty \left(\frac{d\eta}{dt}\right) dt,$$

where the line integral is taken along the integral curve in $\mathfrak{F}$ issuing from (x_0, y_0, x_0', y_0'). If we can show that this integral is a continuous function of (x_0, y_0, x_0', y_0') on $\mathfrak{B}$, the truth of our theorem will follow at once; for, within a neighborhood of P, (x, y, x', y') are continuous (analytic, in fact) in $(x_0, y_0, x_0', y_0', t)$, and hence in $(x_0, y_0, x_0', y_0', \omega[x, y, x', y'])$; so that (x_0, y_0, x_0', y_0') may be expressed in terms of such functions of (x, y, x', y').

Since the integrand is continuous (analytic), we have but to establish the uniform convergence of the integral $\int_0^\infty (d\eta/dt)dt$ within any closed subregion of $\mathfrak{B}$. Writing

$$F(t) = F(x_0, y_0, x_0', y_0', t) = \int_0^t \left(\frac{d\eta}{dt}\right) dt,$$

we must prove that, if a constant $\epsilon > 0$ be assigned, there is a T independent of (x_0, y_0, x_0', y_0') such that $|F(t_2) - F(t_1)| \leqq \epsilon$ for all $|t_1| > T$, $|t_2| > T$. We have, to begin with,

$$F(t_2) - F(t_1) = \int_{t_1}^{t_2} \left(\frac{d\eta}{dt}\right) dt.$$

Further, by (1), and I, (4),

$$\frac{d\eta}{dt} = y \frac{\partial \Pi}{\partial x} - x \frac{\partial \Pi}{\partial y} = -\frac{\partial \Pi}{\partial \theta}.$$

A little computation shows that there exists a positive constant M such that $|\partial \Pi/\partial \theta| \leqq M/r^2$ (indeed, $\leqq M/r^3$). Further, by Chapter I, formula (4) and $G = r^2 d\theta/dt$, we have

$$\lim_{t \to +\infty} \frac{dr}{dt} = (2\eta_\infty)^{1/2},$$

the limit being approached from above. Hence, a constant $h>0$ exists such that, within $\mathfrak{F}$, $dt \leqq h dr$. Hence, if $t_2 \geqq t_1$,

$$|F(t_2)| - |F(t_1)| \leqq Mh \int_{r_2}^{r_1} \frac{dr}{r^2} = Mh\left(\frac{1}{r_1} - \frac{1}{r_2}\right),$$

and this is sufficient to establish the uniform convergence.

THEOREM 3. *For any given value of the constant of Jacobi, there exist infinitely many orbits going to infinity with any assigned value of* $\eta_\infty > 0$. *A similar theorem is true for orbits which come from infinity.*

By Chapter I, Theorem 4, there exist orbits going to infinity with η_∞ smaller than, or greater than, any assigned positive quantity. Since all orbits whose existence is given by that theorem have $\eta_\infty > 0$, they belong to $\mathfrak{M}'$. From the continuity of η_∞, and the connected nature of $\mathfrak{M}'$, our present theorem results.

The existence of orbits going to, or coming from, infinity with $\eta_\infty = 0$ will be established in Chapter IV, §2, in the case where there exists a surface of section.

CHAPTER III. THE SURFACE OF SECTION

1. **Preliminary remarks.** Consider a 3-dimensional manifold, the points of which form a closed set, and which, at each point, is analytic and without singularity. We shall call this the *manifold of states of motion*, a terminology explained by the fact that we shall think of its points as being in one-to-one correspondence with the *states of motion* (i.e., position-and-velocity) of a dynamical system. Let there be defined a set of analytic curves in the manifold, just one through each point, and of such a nature that in a suitable neighborhood $\mathfrak{A}$ of each point they appear as the integral curves of the differential system

$$\frac{dx}{dt} = X(x,y,z), \qquad \frac{dy}{dt} = Y(x,y,z), \qquad \frac{dz}{dt} = Z(x,y,z),$$

where x, y, z are Gaussian coördinates figuring in the analytic representation of the manifold of states of motion in $\mathfrak{A}$; X, Y, Z are single-valued, analytic, and nowhere all zero within this neighborhood. We shall say that we have a *steady flow* in the manifold, and speak of the curves as the *stream lines*, the parameter t as the time. It is understood that this differential system comes to us from the equations of motion of the dynamical system in question; thus the stream lines, and the trajectories of the latter, correspond. We shall admit that there may be a finite number of exceptional points, and a finite number of exceptional curves (which are simple, closed,

and analytic), at which these properties of the stream lines do not hold. These will be referred to as the *singularities* of the flow.

By *surface of section** we shall mean a connected surface, the points of which form a closed set, having the following properties:

(a) The surface is analytic and without singularities at each of its points.

(b) The boundaries (if any) are either stream lines, or singularities of the flow; and every singularity appears upon the boundary.

(c) Every stream line cuts the surface at least once within every interval of time of length τ, where τ is the same for all stream lines.

(d) No stream line is tangent to the interior of the surface, and the angle made with the surface by the stream lines either does not approach zero, or else is at most of the first order in the distance from the boundaries.

It is a familiar fact that the restricted problem of three bodies may be divided into various different cases, depending on the form of the ovals of zero velocity, and on the particular region in which the motion takes place. The manifold of states of motion which it is important to consider in the present connection is any one of the connected manifolds of the points (x, y, x', y') which correspond to a given value of the constant of Jacobi. They are defined analytically by the equation

$$x'^2 + y'^2 = 2\Omega - C$$

(cf. Chapter I, (3)). In the case where the infinitesimal body moves inside a closed oval containing only one of the massive bodies S, J, Birkhoff† has shown that a ring-shaped surface of section exists, provided μ be taken sufficiently small, and the constant of Jacobi $C > 32^{1/3}$. It will be shown in this chapter that, with precisely the same restrictions for μ and C, a ring-shaped surface of section exists in the case of motion *outside* of an oval of zero velocity containing both S and J. We shall use the method employed so frequently

---

* Cf. the developments of this idea in the following:

H. Poincaré, *Les Méthodes Nouvelles de la Mécanique Céleste*, Paris, Gauthier-Villars, 1899, vol. III, pp. 196–200, 372–381.

H. Poincaré, *Sur un théorème de géométrie*, Rendiconti del Circolo Matematico di Palermo, vol. 33 (1912), pp. 375–407.

G. D. Birkhoff, *The restricted problem of three bodies*, Rendiconti del Circolo Matematico di Palermo, vol. 39 (1915), pp. 1–70.

G. D. Birkhoff, *Dynamical systems with two degrees of freedom*, these Transactions, vol. 18 (1917), pp. 199–300.

G. D. Birkhoff, *Surface transformations and their dynamical applications*, Acta Mathematica, vol. 43 (1920), pp. 1–119.

† G. D. Birkhoff, *The restricted problem of three bodies*, Rendiconti del Circolo Matematico di Palermo, vol. 39 (1915), §13.

by Poincaré* and Birkhoff,† which consists in finding an explicit representation of the surface of section in the integrable case $\mu=0$, and generalizing by continuity and analytic continuation of periodic orbits to the case $\mu>0$. It may be remarked that the representation which we shall obtain may be used, with slight modifications, in the case considered by Birkhoff.

2. **The surface of section for $\mu=0$.** When $\mu=0$, our problem reduces to the problem of two bodies, referred to coördinate axes through the central body which rotate with unit angular velocity in the positive direction. The orbits, therefore, are conic sections, with a focus at the origin, rotating with respect to the xy-axes with unit angular velocity in the negative direction.

The manifold of states of motion which we shall consider is given by

$$(1) \qquad v^2 = r^2 + \frac{2}{r} - C, \qquad r = (x^2 + y^2)^{1/2}, \qquad v^2 = x'^2 + y'^2.$$

(Cf. Chapter I, (3).) For a given value of $C>3$, the ovals of zero velocity consist of two circles with centers at the origin and radii $r'<1$ and $r''>1$. Motion is possible within the smaller, or without the larger circle. The latter case is the one we shall consider.

In order to examine the nature of the integral curves in this manifold, we shall transform to a new set of variables, which may be regarded as *elements* of the orbits. For this purpose it is not possible to use the ordinary elliptic elements, for we are in the presence of parabolic and hyperbolic orbits, for which the elliptic semi-major axis "a" becomes infinite and imaginary, respectively. It will be seen that the variables denoted below by (e, ϑ, w) are suitable for our purpose.

It will be convenient to introduce the following quantities, together with certain relations between them, familiar in celestial mechanics. Let

e = eccentricity of the orbit;

ϑ = longitude of the perihelion, measured from the $+x$-axis, and reduced to the interval $0 \leqq \vartheta < 2\pi$;

w = true anomaly of the particle, reduced to the interval $-\pi \leqq w < \pi$;

$$a = \begin{cases} \text{elliptic semi-major axis} & \text{when } e<1 \\ \text{not defined} & \text{when } e=1 \\ \text{hyperbolic semi-major axis} & \text{when } e>1; \end{cases}$$

* H. Poincaré, *Les Méthodes Nouvelles de la Mécanique Céleste*, Paris, Gauthier-Villars, 1899, pp. 196–200.

H. Poincaré, *Sur un théorème de géométrie*, Rendiconti del Circolo Matematico di Palermo, vol. 33 (1915), pp. 375–407.

† G. D. Birkhoff, *The restricted problem of three bodies*, Rendiconti del Circolo Matematico di Palermo, vol. 39 (1915), §13.

η and G have the same meaning as previously. They are related by Chapter I, (3′):

$$2(G - \eta) = C.$$

The polar equation of the orbit as given in celestial mechanics is

$$r = \frac{G^2}{1 + e \cos w}. \tag{2}$$

Comparing this with the corresponding equation given in geometry,

$$r = \frac{\pm\, a(1 - e^2)}{1 + e \cos w}, \qquad e \gtrless 1,$$

we derive

$$G^2 = \begin{cases} + a(1 - e^2) & \text{when } e < 1 \\ - a(1 - e^2) & \text{when } e > 1. \end{cases} \tag{3}$$

At perihelion, $dr/dt = 0$, $w = 0$, and

$$r = r_0 = \begin{cases} + a(1 - e) & \text{when } e < 1 \\ - a(1 - e) & \text{when } e > 1 ; \end{cases}$$

also

$$\eta = \frac{1}{2}V^2 - \frac{1}{r_0} = \frac{1}{2}\frac{G^2}{r_0^2} - \frac{1}{r_0}.$$

By the use of these two formulas and of (3) we get

$$\eta = \begin{cases} -\dfrac{1}{2a} & \text{when } e < 1 \\ 0 & \text{when } e = 1 \\ +\dfrac{1}{2a} & \text{when } e > 1. \end{cases} \tag{4}$$

The middle formula follows from the fact that in parabolic motion the velocity V, and hence η, are zero at infinity. Since (4) is a relation between constants, it holds for a general value of w. Excluding for the moment the case $e = 1$, we get, on eliminating a from (3) and (4):

$$2\eta G^2 = e^2 - 1.$$

Using this to eliminate G from $2(G - \eta) = C$, we obtain

$$f(\eta) \equiv \eta^3 + C\eta^2 + \frac{C^2}{4}\eta + \frac{1 - e^2}{2} = 0. \tag{5}$$

Although deduced for $e \neq 1$, both these formulas follow at once from (4) when $e=1$.

We are now able to draw some conclusions. Since attention is being restricted to motion outside the circle of radius $r''>1$, it is seen that, when $e<1$, $a \geqq r''>1$. Hence, by (4),

$$\eta \geqq -\frac{1}{2r''} > -\frac{1}{2}. \tag{6}$$

From this, from $2(G-\eta)=C$, and from the fact that $C>3$, it follows that

$$G \geqq \frac{1}{2}\left(C - \frac{1}{r''}\right) > 1.$$

And when $e \geqq 1$, clearly $G>0$. Hence *there are no retrograde orbits.**

In order to examine the manner in which η depends on e through equation (5), let the curve $\xi=f(\eta)$ be plotted. It is seen that $f'(\eta)=0$ has the two roots

$$\eta = -\frac{C}{6}, \qquad \text{when } f\left(-\frac{C}{6}\right) = \frac{1-e^2}{2} - \frac{1}{2}\left(\frac{C}{3}\right)^3;$$

$$\eta = -\frac{C}{2}, \qquad \text{when } f\left(-\frac{C}{2}\right) = \frac{1-e^2}{2}.$$

Further, $f(-\infty)=-\infty$, and $f(+\infty)=+\infty$. Hence the curve has the character shown in Fig. 3.

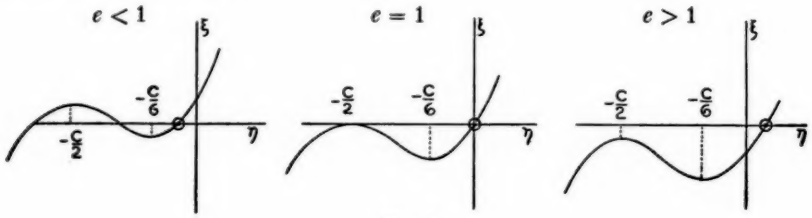

Fig. 3

It follows that for each value of e there is just one value of η satisfying both (5) and (6); for by substitution, $f(-\frac{1}{2})<0$. Now regard η as a function of e defined by (5): $\eta=F(e)$. Differentiating, we get

$$F'(e) = \frac{d\eta}{de} = \frac{e}{f'(\eta)},$$

* This theorem, which is still true when $\mu>0$ and small, i.e., the case we are going to consider, can be proved for a general value of μ, provided C is not too small. For then it is easy to show by Chapter I, (7), that $G>0$.

and this is seen from the plot to be positive whenever $e>0$. This and (5) show that *as e increases from 0 to $+\infty$, the unique corresponding $\eta=F(e)$ increases from $\eta_0>-\frac{1}{2}$ to $+\infty$.*

Let us obtain the parametric representation of the manifold of states of motion (1) in the curvilinear coördinates (e, ϑ, w). Differentiation of (2) along an integral curve gives

$$\frac{dr}{dt}=\frac{G^2e\sin w}{(1+e\cos w)^2}\frac{dw}{dt};$$

with this, (2), and $G=r^2d\theta/dt$, we can eliminate $d\theta/dt$ and dr/dt from the integral of energy. Use of the formula $2\eta G^2=e^2-1$ proved above reduces the result to

$$\frac{dw}{dt}=\frac{(1+e\cos w)^2}{G^3}.$$

Thus, the stream lines in the $e\vartheta w$-space are determined by the system

$$(7)\qquad \begin{aligned}\frac{de}{dt}&=0,\\ \frac{d\vartheta}{dt}&=-1,\\ \frac{dw}{dt}&=\frac{(1+e\cos w)^2}{G^3},\end{aligned}$$

where it is assumed that G is replaced by its value in e from (5):

$$G=\eta+\frac{C}{2}=F(e)+\frac{C}{2}.$$

We have but to introduce the polar coördinates (r, α) in the rotating xy-plane, when the required uniformizing equations are found to be

$$(8)\qquad \begin{aligned}r&=\frac{G^2}{1+e\cos w},\\ \alpha&=\vartheta+w,\\ r'&=\frac{e\sin w}{G},\\ \alpha'&=-1+\frac{(1+e\cos w)^2}{G^3},\end{aligned}$$

where as before $G=F(e)+C/2$.

Through any point (x, y, x', y'), or (r, α, r', α'), of the manifold of states of motion (1) there passes a unique stream line, and hence there corresponds a unique set of values (e, ϑ, w) in the intervals considered, provided $e>0$. Conversely, it is seen from (8) and (5) that to every set (e, ϑ, w) such that $1+e \cos w>0$, $e>0$, there corresponds a unique set of (r, α, r', α') and of (x, y, x', y') which are shown to belong to (1).

In order to obtain an adequate representation of the flow, we introduce cylindrical polar coördinates, taking e as radius-vector, ϑ as angle, and w as altitude. Our manifold of states of motion in the resulting $e\vartheta w$-space appears as the solid of revolution formed by revolving the portion of the half-plane $\vartheta=0$ within $ABcc'B'A'$ (shaded in Fig. 4) about the w-axis. This is all seen from the geometric meaning of our variables. The surfaces traced by Bc and $B'c'$ have the equation $1+e \cos w=0$.

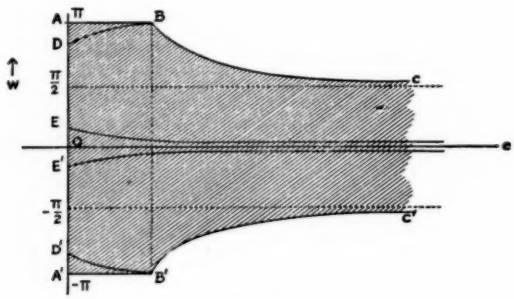

Fig. 4

The single infinitude of states of motion in the circular orbit is appropriately represented by the segment AA' of the w-axis. This leads us to define the variable w on such an orbit: chose an arbitrary point upon it, stationary in the fixed space, and let w be the angle measured from this point to the particle.

We shall now make the convention that *all pairs of points* $(e, \vartheta, -\pi)$ *and* (e, ϑ, π) *shall be regarded as identical.* With this it at once becomes clear from the geometric meaning of the variables that *the manifold of states of motion* (1) *of the* $xyx'y'$*-space is in continuous one-to-one correspondence with this open torus-like region of the* $e\vartheta w$*-space, the singularity at* ∞ *corresponding to the boundary of the torus.*

The stream lines in the $e\vartheta w$-space are given by the finite equations

$$(9) \qquad e = e_0, \qquad \vartheta = -t + \vartheta_0, \qquad w = w(t) \qquad [w'(t)>0].$$

For $e_0<1$, they are curves on the cylinder $e=e_0$, winding steadily upwards, but with variable pitch. They vary from the circular motions $e_0=0$ to the parabolic motions $e_0=1$. As they approach the latter, the pitch at $w=\pm\pi$ approaches zero. The parabolic motions wind infinitely often about the w-axis, approaching the plane $w=\pi$ as $t\to+\infty$, and $w=-\pi$ as $t\to-\infty$. The hyperbolic motions are similar in character to the parabolic, only as $t\to\pm\infty$ they approach the surfaces traced by the revolution of the curves Bc and $B'c'$, corresponding to approach to infinity $(1+e\cos w=0)$.

Consider the portion of the plane $\vartheta=0$ which is shown in the shaded region $ABcA'B'c'$ in Fig. 4. Because of our convention regarding points on AB and $A'B'$, it is of the nature of a ring, bounded by the circular orbit, and by a closed curve $Bcc'B'$ whose neighborhood corresponds to the neighborhood of infinity in the xy-plane. For convenience, we shall refer to these as the "inner" and the "outer" boundaries respectively. It is evident from equations (7) that this ring is cut throughout in the same sense by the stream lines, once by each in the interval of time 2π. The stream lines are never tangent to the interior of the ring, but make an angle which approaches $\pi/2$ at the outer boundary, and is of the first order in the distance from the inner boundary. The pitch of these lines approaches the constant value of $-a_0^{-3/2}$ $(a_0=a|_{e=0})$ at the latter boundary. Thus the transformation of the ring into itself, which the dynamical problem defines, may be extended by continuity to the inner boundary: it advances all points of this a distance of $a_0^{-3/2}$ upwards. It is clear that *the ring is a surface of section for the problem in the $e\vartheta w$-space.*

3. **The surface of section for $\mu>0$.** It was proved by Birkhoff* that the circular orbits in the case $\mu=0$ admit of a unique analytic continuation into symmetric periodic orbits without double points, for a sufficiently small choice of $\mu>0$, provided $C>32^{1/3}$. Although stated for motion inside the oval containing one of the massive bodies, it may be seen that the proof holds verbatim for the exterior motion we are considering.

We wish to perform a certain transformation on the $xyx'y'$-space, in order to carry the manifold of states of motion for $\mu>0$ into that for $\mu=0$. Introduce the polar coördinates (r, α) in the xy-plane, and (v, β) in the $x'y'$-plane. Let $r=r''$ and $r=r_0$ be the equation of the oval of zero velocity and the circular orbit respectively when $\mu=0$; and $r=\Phi(\alpha)$, $r=\Psi(\alpha)$, those of the oval and the analytic continuation when $\mu>0$. Suppose, further, that $\beta=g(\alpha)$ and $\alpha=h(t)$ for the latter. Consider the transformation

* G. D. Birkhoff, *The restricted problem of three bodies*, Rendiconti del Circolo Matematico di Palermo, vol. 39 (1915), §11.

$$\bar{r} = [\Psi(\alpha) - \Phi(\alpha)]^{-1} r^{-2} \{ [\Psi(\alpha) - \Phi(\alpha)] r^3 + [\Psi(\alpha)^2 \{ r_0 - \Psi(\alpha) \} - \Phi(\alpha)^2 \{ r'' - \Phi(\alpha) \}] r + \Psi(\alpha)\Phi(\alpha)^2 [r'' - \Phi(\alpha)] - \Phi(\alpha)\Psi(\alpha)^2 [r_0 - \Psi(\alpha)] \},$$

$$(10) \qquad \bar{\alpha} = \alpha + [AH(\alpha) - \alpha]\left(\frac{r_0}{\bar{r}}\right)^2, \qquad \left(A = \frac{1}{\frac{C}{2} - F(0)} - 1 \text{ ; cf. (8)}\right),$$

$$\bar{v} = v\left(\frac{P}{Q}\right)^{1/2},$$

$$\bar{\beta} = \beta + \left(\alpha + \frac{\pi}{2} - g(\alpha)\right)\left(\frac{r_0}{\bar{r}}\right)^2.$$

Here H is the inverse of h; it is single-valued and analytic, since h is, and $dh/dt > 0$; P is the quantity $2\Omega - C$ for $\mu = 0$ formed for the point $(\bar{r}, \bar{\alpha})$ (expressed in terms of (r, α) by means of the first two equations of (10)), while Q is the value of $2\Omega - C$ for the given $\mu > 0$ formed for the point (r, α). Since the ovals correspond in virtue of the first two equations of (10), it follows that P and Q are *analytically equivalent** on the oval; i.e., $P \equiv QR$, where R is analytic and distinct from zero at all points of the oval. Hence, $(P/Q)^{1/2}$ can be defined in its removable singularities so that it is analytic for all the points considered. Finally, it will be seen that the jacobian of the transformation approaches unity as $r \to \infty$; and since it is unity when $\mu = 0$ ((10) reducing to the identity), and is analytic, we may clearly take $\mu > 0$ so small that the jacobian is greater than a certain positive constant throughout the region considered. It is easy to see now that, for $\mu > 0$ sufficiently small, (10) has the following properties:

1. It is an analytic correspondence between the manifold of states of motion for $\mu = 0$ and that for the given $\mu > 0$, varying analytically with μ, and reducing to the identity when $\mu = 0$; the dashed letters refer to points of the former.

2. It carries the circular orbit for $\mu = 0$ into its analytic continuation for $\mu > 0$; and the transformed point describes the circle $\bar{r} = r_0$ at the same uniform rate as when $\mu = 0$.

3. The displacement which it produces in the points is uniformly bounded, and approaches zero uniformly as $r \to \infty$.

---

* Cf. W. F. Osgood, *Lehrbuch der Funktionentheorie*, Leipzig, B. G. Teubner, 1924, vol. II, pp. 78, 87.

4. It is *symmetric*; i.e., if $(\bar{x}, \bar{y}, \bar{x}', \bar{y}')$ corresponds to (x, y, x', y'), then $(\bar{x}, -\bar{y}, -\bar{x}', \bar{y}')$ will correspond to $(x, -y, -x', y')$.

The last property results from the fact that the equations of motion, Chapter I, (2), are left unchanged when (x, y, t) are replaced by $(x, -y, -t)$, and the consequent symmetric nature of all the elements which enter into (10).

Consider now the transformation of the manifold $v^2 = 2\Omega - C$ into the solid ring in the $e\vartheta w$-space of Fig. 4 obtained on the successive application of (10) and (8). In order to see what is going on, we shall invert the transformation (8). It is convenient to replace the cylindrical coördinates (e, ϑ, w) by the rectangular coördinates (p, q, w): $p = e\cos\vartheta$, $q = e\sin\vartheta$. The equations (8) yield

$$p = \cos\alpha\cdot\left(\frac{G^2}{r} - 1\right) + \sin\alpha\cdot Gr' = \frac{x}{r}\left(\frac{G^2}{r} - 1\right) + \frac{y}{r}Gr',$$

$$(8') \qquad q = \sin\alpha\cdot\left(\frac{G^2}{r} - 1\right) - \cos\alpha\cdot Gr' = \frac{y}{r}\left(\frac{G^2}{r} - 1\right) - \frac{x}{r}Gr',$$

$$w = \arctan\frac{Gr'}{\frac{G^2}{r} - 1}, \qquad G = r^2(\alpha' + 1).$$

It is understood that the points (r, α, r', α'), or (x, y, x', y'), belong to the given manifold of states of motion. These equations are analytic, except for the last when $G^2 = r$. The periodic orbit $r = \Psi(\alpha)$ is carried into the axis $e = 0$, and all the stream lines go into the curves

$$e = f_1(t, \mu), \qquad \vartheta = f_2(t, \mu), \qquad w = f_3(t, \mu),$$

which are analytic in t and μ for μ sufficiently small, and reduce to (9) when $\mu = 0$. We wish to show that the surface $\vartheta = 0$ is related to the stream lines in a manner quite similar to that in the case $\mu = 0$ of the last section.

Let us visualise what happens as μ is increased continuously from zero: the stream lines vary analytically with μ, and will continue to cut the ring in the requisite manner as long as $d\vartheta/dt < \text{const.} < 0$, and dw/dt remains finite.

The differential equations of motion in the $e\vartheta w$-space may be obtained by differentiating the relations $e = (p^2 + q^2)^{1/2}$, $\vartheta = \tan^{-1}(q/p)$, and making use of the equations (8'). We have

$$\frac{de}{dt} = \cos\vartheta\frac{dp}{dt} + \sin\vartheta\frac{dq}{dt} = \cos(\alpha-\vartheta)\frac{d}{dt}\left(\frac{G^2}{r}\right) + \sin(\alpha-\vartheta)\frac{d(Gr')}{dt},$$

$$\frac{d\vartheta}{dt} = \frac{1}{e}\left[\cos\vartheta\frac{dq}{dt} - \sin\vartheta\frac{dp}{dt}\right]$$

$$= \alpha' + \frac{1}{e}\left[\sin(\alpha-\vartheta)\frac{d}{dt}\left(\frac{G^2}{r}\right) - \cos(\alpha-\vartheta)\frac{d(Gr')}{dt}\right],$$

$$\frac{dw}{dt} = \frac{1}{e}\left[\cos w\frac{d(Gr')}{dt} - \sin w\frac{d}{dt}\left(\frac{G^2}{r}\right)\right].$$

A further use of equations (8) reduces these to

$$\frac{de}{dt} = \cos w\frac{d}{dt}\left(\frac{G^2}{r}\right) + \sin w\frac{d(Gr')}{dt} = F_1,$$

$$\frac{d\vartheta}{dt} = -1 + \frac{(1+e\cos w)^2}{G^3} + \frac{1}{e}\left[\sin w\frac{d}{dt}\left(\frac{G^2}{r}\right) - \cos w\frac{d(Gr')}{dt}\right] = F_2,$$

$$\frac{dw}{dt} = -\frac{1}{e}\left[\sin w\frac{d}{dt}\left(\frac{G^2}{r}\right) - \cos w\frac{d(Gr')}{dt}\right] = F_3.$$

I say that the F's are analytic in general, and continuous in all the variables for all the points considered. The only points where trouble might arise are on the axis $e=0$. But if we set $G=r^2(\alpha'+1)$, and perform the differentiations, we shall see that the $(1/e)$-terms all occur in the form r'/e, r''/e, and α''/e. The first, by equation (8), has the value $(\sin w)/G$. Further, r'' and α'' vanish when $e=0$; for then we have uniform circular motion (cf. property 2 of (10)). But r'' and α'' are analytic in (r, α, r', α') (cf. the polar equations of motion; this property is not changed by (10)); hence they vanish at least as fast as an expression of the form $K_1(r-r_0)+K_2(\alpha-\alpha_0)+K_3(r'-r_0')+K_4(\alpha'-\alpha_0')$, $K_i=$const., $(r_0, \alpha_0, r_0', \alpha_0')$ being a point on the orbit $e=0$. Here we may assume $K_2=0$, since $r''=\alpha''=0$ when $r=r_0$, $r'=r_0'$, $\alpha'=\alpha_0'$, $\alpha\neq\alpha_0$. And from (8) it appears that $r-r_0$, $r'-r_0'$, $\alpha'-\alpha_0'$ are of the first order in e. Hence r''/e and α''/e are continuous, and the same is true of the above equations.

When $\mu=0$, $F_2=-1$; also, F_2 approaches -1 at the outer boundary of the ring, for the corresponding motion approaches two-body motion uniformly at infinity, and by the property 3, the deformation (10) reduces uniformly to zero as $r\to\infty$. It follows that $\mu>0$ may be taken so small that $d\vartheta/dt<$const.<0 throughout. Furthermore, F_2 is everywhere finite when $\mu=0$, and since it is finite in the neighborhood of the outer boundary when $\mu>0$, we may take $\mu>0$ so small that F_3 is finite throughout.

The transformation established upon the ring by the stream lines advances points of the inner boundary by an amount which is a continuous function of μ, reducing to $a_0^{-3/2}$ when $\mu=0$; hence, for a sufficiently small μ, this amount remains greater than a constant, ω. Similarly, the time interval between consecutive intersections is analytic in μ, reducing to 2π when $\mu=0$, and at the outer boundary; hence, for a sufficiently small μ, this interval remains less than a constant, τ. Thus we have proved the fundamental theorem:

THEOREM 1. *When $C>32^{1/3}$, it is possible to take a sufficiently small $\mu_0>0$ so that, for each value $0\leqq\mu\leqq\mu_0$, there exists a ring-shaped surface of section for the problem of exterior motion we are considering. The inner boundary is constituted by a closed stream line which corresponds to a symmetric periodic orbit without double points. The outer boundary corresponds to the infinite region in the xy-plane. The transformation T established on the ring by the flow advances all points of the inner boundary by an amount greater than a constant $\omega>0$. There are points arbitrarily near to the outer boundary which never make a complete circuit of the ring however often T is applied. Replacing (x, y, x', y') by $(x, -y, -x', y')$,—or, what is equivalent, (e, ϑ, w) by $(e, -\vartheta, -w)$,—carries the ring into itself.*

The statement about the rotation in the neighborhood of the outer boundary follows from the existence of half-orbits which reach to infinity without apsides; cf. the interpretation of Fig. 4.

THEOREM 2. *Neither of the two boundaries of the image of the ring in the $xyx'y'$-space can be shrunk continuously to a point without either passing through the other boundary, or departing from the manifold of states of motion $v^2=2\Omega-C$.*

It was shown by Birkhoff* that if the variable $\phi=\alpha=\arctan(y'/x')$ be introduced, the manifold of states of motion $v^2=2\Omega-C$ in the $xyx'y'$-space is in one-to-one continuous correspondence with points (x, y, ϕ), $0\leqq\phi<2\pi$, except that a single point (x_0, y_0, x_0', y_0') on the oval of zero velocity corresponds to the single infinitude of points (x_0, y_0, ϕ), $0\leqq\phi<2\pi$. This leads us to the following geometric representation.†

First, deform the region of motion of the xy-plane into the exterior of a circle K in that plane, whose center is at the origin. Let (x, y) correspond to

* G. D. Birkhoff, *The restricted problem of three bodies*, Rendiconti del Circolo Matematico di Palermo, vol. 39 (1915), p. 21.

† This representation was first applied (incorrectly) by H. Poincaré to the case of motion inside an oval containing one of the massive bodies. Cf. *Sur un théorème de géométrie*, Rendiconti del Circolo Matematico di Palermo, vol. 33 (1912), pp. 375–407.

$(\bar{x}, \bar{y})$ in this deformation. Next, construct all possible circles as follows: (1) the planes of the circles are normal to the xy-plane; (2) their centers are in this plane; (3) their two points of intersection with the xy-plane are mutual inverts in the circle K. Finally, let the point (x, y, x', y') of the manifold of states of motion be represented by the point upon the circle through $(\bar{x}, \bar{y})$ such that the arc of this circle subtended by $(\bar{x}, \bar{y})$ and the representative point, measured upwards from $(\bar{x}, \bar{y})$, is of angular measure ϕ.

It is clear that the correspondence is everywhere one-to-one and continuous, and that the points corresponding to the singularity at infinity lie on the z-axis in the representative space. This axis clearly threads the periodic orbit, which proves our theorem. We are, of course, using the space of inversion.

We note in passing that the above representation gives an obvious proof of Birkhoff's theorem to the effect that the manifold may be deformed into the space of inversion from which a simple closed curve, corresponding to infinity, has been removed.

4. **Two properties of the transformation.*** Consider the differential system of the first order, Chapter II, (5), which governs the flow in the $xyx'y'$-space. The flow is clearly *incompressible*, i.e., it admits the integral invariant $\iiiint dx\, dy\, dx'\, dy'$. For the "equation of continuity" is obviously satisfied. It follows, by reasoning which has been used so often in dynamics that it is unnecessary to repeat it here, that the flow in the manifold of states of motion $v^2 = 2\Omega - C$ has a positive volume integral invariant, and that *the transformation T on the surface of section has a positive area integral invariant.* The latter approaches zero at the boundary corresponding to the periodic orbit.

Another property of importance is the following:

The transformation T is the product of two transformations, R and U, each of period 2, where R consists in the substitution of $(x, -y, -x', y')$ for (x, y, x', y'),—or of $(e, -\vartheta, -w)$ for (e, ϑ, w).

Consider the two points A and B of the surface of section, where $B = T(A)$. By definition of T, B is the next intersection with the surface of the stream line through A. Since the equations of motion are unchanged when x, y, and t are replaced by x, $-y$, and $-t$, it is seen that the transform of the

* This section is but a paraphrase of the same facts established by Birkhoff for interior motion. Cf. *The restricted problem of three bodies*, Rendiconti del Circolo Matematico di Palermo, vol. 39 (1915), §14.

stream line $\{AB\}$ by R passes through the points $R(A)$ and $R(B)$, but in the opposite order in time. Therefore

$$R(A) = TR(B) = TRT(A).$$

Since $R^2=I$, the identity, we have $A=RT\ RT(A)$, i.e. $RTRT=I$. Hence if we set $U=RT$, we have $T=RU$, and $R^2=U^2=I$. This fact is true by continuity on the inner boundary.

CHAPTER IV. THE DISTRIBUTION OF PERIODIC AND UNSTABLE MOTIONS

1. **Symmetric periodic orbits.** Throughout this chapter we shall consider the motion of the particle outside the oval of zero velocity containing S and J, and shall assume C and μ to be so restricted that all the results of Chapter III regarding the existence of the surface of section and the transformation T are in force.

In view of the properties of T given by Chapter III, Theorem 1, we have a right to conclude, by Poincaré's Geometric Theorem,* that there exist infinitely many periodic orbits for a given value of C. But the symmetric properties of the transformation (see Chapter III, §4) make it possible to arrive at the same result, and at the same time to obtain certain important properties of those of the periodic orbits which are symmetric in the x-axis. This state of affairs was taken advantage of by Birkhoff, in the case of interior motion.† The present section is but the most obvious extension of his methods and results to the case of exterior motion we are considering.

We have seen that the differential equations of motion, Chapter I, (2), are unchanged if (x, y, t) be replaced by $(x, -y, -t)$. From this it follows that if we have an orbit which crosses the x-axis twice at right angles, this orbit is a symmetric periodic orbit, composed of the portion lying between these two intersections, and its symmetric image in the x-axis. Hence no orbit can have more than two such crossings. Further, every periodic orbit which is symmetric in the x-axis crosses it at least twice at right angles. For if at $t=t_1$ and $t=t_2$ we have a pair of successive symmetric crossings at the same point, there will be a right angle crossing when $t=t_3=(t_1+t_2)/2$. If t_4 is the next time the particle crosses at right angles

---

* H. Poincaré, *Sur un théorème de géométrie*, Rendiconti del Circolo Matematico di Palermo, vol. 33 (1912), pp. 375–407.

G. D. Birkhoff, *Dynamical systems with two degrees of freedom*, these Transactions, vol. 18 (1917), §34.

† G. D. Birkhoff, *The restricted problem of three bodies*, Rendiconti del Circolo Matematico di Palermo, vol. 39 (1915), §20.

at this same point, it will cross at right angles, and at a different point, when $t=(t_3+t_4)/2$. Thus our statement is proved.

Right angle crossings of the x-axis may be divided into four classes, according to whether they lie to the right or the left of the oval of zero velocity, and correspond to a relative maximum or minimum distance of the particle from the origin. We shall distinguish

class I, lower apsis at the right of the oval $(\vartheta=0,\ w=0)$;
class II, higher " " " " " " " $(\vartheta=\pi,\ w=-\pi)$;
class III, lower " " " left " " " $(\vartheta=\pi,\ w=0)$;
class IV, higher " " " " " " " $(\vartheta=0,\ w=-\pi)$.

We have indicated the points in the $e\vartheta w$-space corresponding with the states of motion of the respective crossings; the values given are obvious in the case $\mu=0$, and their truth in the general case follows from the symmetric character (property 4) of the transformation (10) of Chapter III.

Consider the ring in the $e\vartheta w$-space (Fig. 4), and the transformation T established upon it by the flow. The lines Oe and $A'B'$ are the loci of the states of motion at crossings of classes I and IV. We have drawn Ee and $E'e$, the respective loci of the first succeeding and first preceding intersections with the ring, made by those stream lines which pass through points of $\vartheta=\pi$, $w=0$,—i.e., which have crossings of class III. The lines $D'B'$ and DB are related in the same way with points of $\vartheta=\pi$, $w=-\pi$ (crossings of class II). By the symmetry, $E'e$ and DB are the reflections of Ee and $D'B'$ in Oe. When $\mu=0$, the curves have the form indicated, and $OE=A'D'=\frac{1}{2}a_0^{-3/2}<\frac{1}{2}$. This continues to be approximately true when $\mu>0$, and small. The essential feature is that Ee is always below Bc, that $D'B'$ always passes through B', and that $OE\geqq\omega>0$, $A'D'\geqq\omega$. (Cf. Chapter III, Theorem 1.)

From the nature of the transformation T in the neighborhood of the boundaries, it follows that by repeating it a sufficient number of times it is possible to make the images of $A'B'$, $D'B'$, Oe, Ee wind an arbitrary number of times around the ring, and hence yield an arbitrary number of intersections with each of these curves in their initial positions. To such an intersection corresponds an orbit having two right angle crossings of the x-axis, of classes depending upon the intersecting curves. We conclude that *there exist infinitely many symmetric periodic orbits with crossings of assigned classes.*

In order to obtain more detailed information with regard to the periodic orbits, we introduce the characteristic integers α, β, analogous to the integers employed by Birkhoff for the interior case, but distinct from those. Let α

be the number of points of intersection that a closed stream line has with the ring; for such an intersection K, $K=T^{\alpha}(K)$; and this equation cannot hold for any number between zero and α. Let β be the number of times that K circulates around the ring after α applications of T. When $\mu=0$, α is the number of rotations of the ellipse in the xy-plane, and β the number of revolutions which the particle makes in this ellipse; β/α is the mean angular velocity in the fixed space.

THEOREM 1. *If k and l are any two integers such that $0<2\pi l\leqq k\omega<2\pi(l+1)$, there will exist a symmetric periodic orbit having crossings of the same class, and with $\alpha=2k$, $\beta=2l$, or any equal submultiples of these.*

After k applications of T, each of the lines $A'B'$, $D'B'$, Oe, Ee has its outer boundary point below Bc, while its point on the inner boundary is advanced by at least $k\omega$. Hence there are on the curve a pair of points, K and L, such that $L=T^k(K)$, the k images of K having gone around the ring a number of times $l\leqq k\omega/(2\pi)$. The points K, L will determine a symmetric periodic orbit with $\beta=2l$ (or a submultiple). It remains to find the corresponding α. Since $T=RU$ (Chapter III, §4), we have

$$K = T^{-k}(L) = (UR)^k(L) = U(RU)^{k-1}R(L),$$

therefore

$$R(K) = T^k[R(L)].$$

When K and L are on $A'B'$ or Oe,

$$R(K) = K, \qquad R(L) = L \qquad (\text{for } A'B' \equiv AB),$$

therefore

$$K = T^k(L).$$

When K and L are on $D'B'$ or Ee, $R(K)$ and $R(L)$ are on DB or $E'e$, therefore

$$R(K) = T^{-1}(K), \qquad R(L) = T^{-1}(L),$$

whence

$$T^{-1}(K) = T^{k-1}(L), \qquad \text{and} \qquad K = T^k(L),$$

as before. Hence, in all cases, $K=T^{2k}(K)$; therefore $\alpha=2k$ (or a submultiple).

THEOREM 2. *If k and l are two integers such that $0<(2l+1)\pi\leqq k\omega<(2l+2)\pi$ there will exist a symmetric periodic orbit having crossings of classes* I *and* III, *or* II *and* IV, *and with $\alpha=2k+1$, $\beta=2l$, or equal submultiples of these.*

After k applications of T, the point E advances by at least $k\omega$. Hence the kth image of Ee cuts Oe in a point L, where $L=T^k(K)$, K on Ee, and the k images of K have circulated around the ring $l \leqq k\omega/(2\pi)-1$ times. This pair of points corresponds to a periodic orbit with crossings of classes I and III, and $\beta=2l$. To find α, we have $R(K)=T^k[R(L)]$ as before. Further, $R(K)=T^{-1}(K)$, and $R(L)=L$. Hence $K=T^{k+1}(L)=T^{2k+1}(K)$, so that $\alpha=2k+1$. Similar reasoning applies to an orbit having crossings of classes II and IV.

THEOREM 3. *If k and l are any integers such that $0<(2l+1)\pi \leqq k\omega<(2l+3)\pi$, there will exist a symmetric periodic orbit having crossings of classes* I *and* IV, *or* II *and* III, *and with $\alpha=2k$, $\beta=2l+1$, or equal submultiples.*

After k applications of T, the point A' advances by at least $k\omega$. If $2\pi l+\pi \leqq k\omega$, the kth image of $A'B'$ will cut Oe in a point L, $L=T^k(K)$, K on $A'B'$. Here as before, $R(K)=T^k[R(L)]$, $R(K)=K$, $R(L)=L$. Hence $K=T^k(L)=T^{2k}(K)$, and $\alpha=2k$. Further, it is seen that $\beta=2l+1$. Similar reasoning is applied in the case of crossings of classes II and III.

THEOREM 4. *If k and l are any integers such that $0<(2l+1)\pi \leqq k\omega<(2l+3)\pi$, there will exist a symmetric periodic orbit having crossings of classes* I *and* II, *or* III *and* IV, *and with $\alpha=2k+1$, $\beta=2l+1$, or equal submultiples.*

The reasoning is just the same as in Theorem 3, only $D'B'$ replaces $A'B'$. Since K is on $D'B'$, we have $R(K)=T^{-1}(K)$. Hence $K=T^{k+1}(L)=T^{2k+1}(K)$ and $\alpha=2k+1$.

2. **The distribution of unstable motions.** In this section we shall study the manner in which orbits going to infinity, orbits coming from infinity, and orbits which both come from and go to infinity are distributed in the manifold of states of motion. For convenience, we shall consider the representation in the $e\vartheta w$-space (Fig. 4), *and shall take μ so small that the transformation T increases the w of each point on the ring by an amount which is positive within the ring, and approaches zero only at $Bcc'B'$.* (T will always have this property near the outer boundary, since there we approach two-body motion. That we may make T behave in this way throughout follows by continuity.)

Consider the surfaces in the manifold of states of motion which are the loci of lower and higher apsides. When $\mu=0$, they are the planes $w=0$ and $w=-\pi$, respectively. For $\mu>0$, they continue to cut the ring in $w=0$ and $w=-\pi$, by the symmetry. Since μ is small, we are sure that the two surfaces have no point in common. Now two consecutive intersections which a stream line may have with one of them are separated by an intersection with

the other. Hence, a stream line issuing from a point of the ring has a higher apsis if and only if it has later intersections with the ring above AB (or below $A'B'$, for t decreasing).

Let Λ denote the set of points on the ring characterized by the fact that the stream line issuing from each one has all its later intersections with the ring approaching the boundary Bc, without approaching B or crossing AB. Such a stream line will correspond to a half orbit which goes to infinity without higher apsides as $t \to +\infty$, and with $\lim_{t\to+\infty}\eta = \eta_\infty > 0$. Theorems 3 and 4 of Chapter I allow us to assert that Λ is not a null set. We will show that it forms a 2-dimensional continuum upon the ring.

Consider the points of the ring for whose correspondents in the $xyx'y'$-space the inequalities

$$\left(\frac{dr}{dt}\right)^2 - \frac{2}{r-1} > 0, \qquad \frac{dr}{dt} > 0, \qquad r > R > 5,$$

are verified. From Chapter I, Theorem 3, and its proof, it follows that they belong to Λ. When $\mu = 0$, they form a band, bounded on one side by the curve Bc; and, from the nature of Chapter I, (8), (10), they will continue to form such a band for $\mu > 0$, provided R be sufficiently large. Now consider any point P of Λ. After a sufficient number, k, of iterations of T, its image $T^k(P)$ will lie within this band, and consequently will be surrounded by a neighborhood of points of Λ. On applying T^{-k} to this neighborhood, it is seen that P is surrounded by a neighborhood of points of Λ. If Q is a second point of Λ, after a sufficient number, l, of repetitions of T, the points $T^l(P)$ and $T^l(Q)$ will both lie in the band, and hence may be connected by a curve of points of Λ. On applying T^{-l} to this curve, it is seen that P and Q may be connected by such a curve. Thus, Λ *forms a continuum of points on the ring, and has* Bc *on its boundary.*

Consider any boundary point of Λ, not on $A'B'$. It is, by definition, a cluster point of points of Λ, and all its images under T possess this property. Hence the boundary points can never rise above AB under iterations of T; otherwise there would be points of Λ which rise above this line,—a contradiction. Since T raises every point by an amount which approaches zero only on Bc and $B'c'$, it follows that our boundary points approach B, i.e., correspond to half orbits which go to infinity without higher apsides. Furthermore, $\lim_{t\to+\infty}\eta = 0$; otherwise the boundary points would belong to Λ, and hence be interior points. We say that a boundary point remains such after an application of T. It is seen that they form a closed set.

Let λ denote the point set defined similarly to Λ, only with $\lim_{t\to+\infty}\eta = 0$, and with the inclusion of B and B'. The points of λ will all approach B as

T is iterated indefinitely. Clearly every point of λ remains such after T. It is easy to see that λ is a closed set. We have seen that λ contains all those points of the boundary of Λ not belonging to Λ.

As λ has infinitely many points on it, it is seen that, *with the restrictions under which we are working, we have completed the Theorem 3 of Chapter* II, *inasmuch as the latter holds for* $\eta_\infty = 0$.

Let Λ' and λ' be defined for $t \to -\infty$ and T^{-1} exactly as Λ and λ were for $t \to +\infty$ and T. Then everything we have said about the latter point sets holds, mutatis mutandis, for the former. Moreover, on account of the symmetrical character of the transformation T, Λ' *and* λ' *are reflections of* Λ *and* λ, *respectively, in the line* Oe. It follows that the boundaries of Λ and Λ' have at least one point in common, which lies on Oe. The fundamental theorem here is the following:

THEOREM 5. *Either* $\Lambda+\lambda$ *and* $\Lambda'+\lambda'$ *coincide, or else* $\Lambda+\lambda$ *has points which descend below* $A'B'$ *on the application of a suitable power of* T^{-1} *and* $\Lambda'+\lambda'$ *has points which ascend above* AB *on the application of a suitable power of* T.

If $\Lambda+\lambda$ and $\Lambda'+\lambda'$ do not coincide, there will, e.g., be a point in the former which does not belong to the latter. This will descend below $A'B'$ on the application of a suitable power of T^{-1}. By the symmetry, the corresponding property is true with relation to AB and T.

When $\mu = 0$, the former alternative of Theorem 5 is the one that holds. λ and λ' consist of the line BB' ($e=1$), which is the boundary of both Λ and Λ'. In general, as long as $\Lambda+\lambda$ and $\Lambda'+\lambda'$ coincide, the distribution of unstable motions among the rest is of this simple type.

Unfortunately, I have not succeeded in finding whether $\Lambda+\lambda$ and $\Lambda'+\lambda'$ continue to coincide when $\mu > 0$. It seems extremely probable from a physical point of view that they do not. For consider a very elongated elliptic orbit in the XY-plane ($\mu = 0$). When the particle is at its higher apsis, let μ take on a positive value. On account of the oscillatory nature of the subsequent force field, the next higher apsis (if any) will in general be distinct from the former. It seems, then, that *at least one orbit exists which has a higher apsis, and yet eventually goes to infinity, in one direction at least.* Such an orbit would insure the distinctness of $\Lambda+\lambda$ and $\Lambda'+\lambda'$. We shall, in the rest of this section, pursue the investigation under the hypothesis that these point sets are distinct.

Only *Theorem* 6 (*below*) *is true without this assumption.*

Let a be a region (or connected point set) which belongs to $\Lambda+\lambda$ but not to $\Lambda'+\lambda'$, and which abuts on Oe; and let b be its reflection in this line.

Then b belongs to $\Lambda'+\lambda'$ but not to $\Lambda+\lambda$, and a and b are separated by points which are in both $\Lambda+\lambda$ and $\Lambda'+\lambda'$. It is clear that points of a divide their orbits into halves, one of which goes to infinity without higher apsides when $t\to+\infty$, while the other does not go to infinity, at least without passing through higher apsides, when $t\to-\infty$. A corresponding property is enjoyed by points of b.

Denote by a_i, b_i the ith images of a and b under T^{-1}. Then all the regions b_i belong to $\Lambda'+\lambda'$, but not to $\Lambda+\lambda$. The corresponding property is not true of the a_i's, except perhaps for the first few. For as in the proof of Theorem 5, there will be points of a certain region a_ν which are below $A'B'$. Suppose that a_ν is the first a_i which has such points, and denote one of the common points of $A'B'$ and a_ν by R_1', its reflection in Oe by R_1. Then it is possible to connect B with R_1' by means of a curve $\{BR_1'\}$ of points of $\Lambda+\lambda$.

Since AB and $A'B'$ are congruent, by our convention with regard to the ring, the regions $a_i(i>\nu)$ will come down into the shaded portion of Fig. 4 again through R_1B. Points of $\Lambda+\lambda$ which have crossed AB k times under iterations of T^{-1} correspond to half orbits which go to infinity as $t\to+\infty$ after passing through k higher apsides. Hence, points which have crossed AB a different number of times cannot coincide. It follows that *there will be points of* $\Lambda+\lambda$ *(e.g., of* a*) which go around the ring (crossing* AB*) an arbitrary number of times as* T^{-1} *is indefinitely applied to the point set.* Thus, the curve $\{BR_1'\}$ of points of $\Lambda+\lambda$ drawn above makes it impossible for those points of $a_{\nu+1}$, $a_{\nu+2}$, $\cdots$ which go below AB to approach $B'c'$; hence, by the nature of T^{-1}, they will descend, eventually past $A'B'$ a second time. By recommencing this reasoning arbitrarily often, drawing each time the curves $\{R_1R_2'\}$, $\{R_2R_3'\}$, $\{R_3R_4'\}$, etc., as we formerly drew $\{AR_1'\}$, we are able to establish the truth of our statement.

The figures $a_{\nu+1}$, $a_{\nu+2}$, $\cdots$ form an infinitude of band-like point sets; as we go arbitrarily far out in this sequence, we come upon bands which spiral around the ring an arbitrary number of times. But since the b's are all above $A'B'$, it follows that between any two a-bands there are bands of points which do not belong either to $\Lambda+\lambda$ or $\Lambda'+\lambda'$, nor to any of their images under iterations of T^{-1}. Points of these separating bands correspond to half orbits which do not go to infinity as $t\to+\infty$; we do not imply, however, that such half orbits remain within any fixed distance from the origin in the xy-plane.

This gives an idea of the complicated nature of the figure which results from $\Lambda+\lambda$ on an indefinite number of applications of T^{-1}. If we reflect in Oe, we obtain the corresponding figure derived from $\Lambda'+\lambda'$ on an indefinite

number of applications of T. These two figures intersect along Oe. Thus there are symmetric orbits which both go to and come from infinity, and have an arbitrary number of higher apsides.

By examining the intersections that are possible between the various types of bands of paragraph-before-last and their reflections, it is likewise seen that there are infinitely many orbits which go to infinity but do not come from infinity, and orbits which come from infinity but do not go to infinity. But, since no intersections of unlike bands of the symmetric figures can occur upon Oe, none of these orbits are symmetric,—a fact which is otherwise obvious.

Consider the infinite succession of ring-shaped regions on the surface of section, contained between the inner boundary and the curves $\{R_iR'_{i+1}\}$ and $\{R'_{i+1}R'_i\}$. As $i\to\infty$, they approach an *inner limiting set*, containing at least the inner boundary. This set is carried into itself by T. Every point of it is on an orbit which remains within a fixed distance from the origin (independent of the orbit) both for $t\to+\infty$ and $t\to-\infty$. (But the converse is not necessarily true.) The set is symmetric in Oe. To points of this set the theory of "Poisson stability" and recurrent motion of Poincaré* and Birkhoff† may be applied.

The inner limiting point set has an area if and only if the periodic orbit constituted by the boundary of the ring is *absolutely stable*, i.e., of such a nature that, whatever $\epsilon>0$ be assigned, there is a neighborhood of points about this orbit all of whose images under T^k and T^{-k} $(k=1, 2, \cdots)$ are within ϵ of the inner boundary.

We shall end this section by a theorem which is entirely independent of the unproved assumption about $\Lambda+\lambda$ and $\Lambda'+\lambda'$. Birkhoff lays down the following definition:‡

"A transformation T will be called *transitive* if, for any pair of points P and Q on S (the surface) nearby points P' and Q' respectively can be found such that $Q'=T^n(P')$.

"A transformation T is *intransitive* in the contrary case."

THEOREM 6. *The transformation T on our surface of section is intransitive.*

If $\Lambda+\lambda$ and $\Lambda'+\lambda'$ coincide, the theorem is obvious, as we see by taking P outside this region, and Q an interior point.

* H. Poincaré, *Les Méthodes Nouvelles de la Mécanique Céleste*, Paris, Gauthier-Villars, 1899, vol. 3, Chapter 26.

† G. D. Birkhoff, *Quelques théorèmes sur le mouvement des systèmes dynamiques*, Bulletin de la Société Mathématique de France, vol. 40 (1912), pp. 305–323.

‡ G. D. Birkhoff, *Surface transformations and their dynamical applications*, Acta Mathematica, vol. 43 (1920), § 61.

If $\Lambda+\lambda$ and $\Lambda'+\lambda'$ do not coincide, take P inside Λ and on Oe. By the symmetry, P will be inside Λ'. Since P is a common point of open regions, it will be surrounded by a neighborhood U of points common to both Λ and Λ'. Let V be the region in which the surface of section is cut by the tube of stream lines issuing from a region of the surface $dr/dt=0, d^2r/dt^2<0$. Since all points of U are on orbits without higher apsides, while those of V are on orbits with higher apsides, it is clear that all the images of the former region under $T^n(n=\pm 1, \pm 2, \cdots)$ are distinct from the latter. If we take Q within V, our theorem is proved.

3. **Some geometrical theorems.** In the corresponding treatment of the interior case, Birkhoff* proved the following theorem:

If Γ be any closed curve (not a closed stream line) in the manifold of states of motion, there will necessarily exist an infinite number of stream lines which pass through a point of Γ twice (or one which does so infinitely many times).

This theorem will still be true in our case under the further hypothesis that *the curve in which the stream lines through Γ next cut the surface of section contains the inner boundary in its interior.* This will insure that this region cuts itself after an application of T, a property which would not otherwise necessarily follow in our case, on account of the infinite extent of the surface of section. Apart from this, the proof goes through without modification.

As corollaries we have the following:

(a) *infinitely many orbits pass twice through any given point of the xy-plane not on the oval of zero velocity*;

(b) *infinitely many orbits pass twice through a point of the oval of zero velocity*;

(c) *infinitely many orbits have two higher apsides at an assigned distance from the origin*; *similarly for lower apsides.*

To show that the theorem applies to (a) and (b), the most convenient way is to examine the position of the states of motion in question in the representation used in the proof of Chapter III, Theorem 2. To prove (c), generalize by continuity from the case of $\mu=0$, in the $e\vartheta w$-space.

---

* G. D. Birkhoff, *The restricted problem of three bodies*, Rendiconti del Circolo Matematico di Palermo, vol. 39 (1915), §21.

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

A CONNECTED AND REGULAR POINT SET WHICH HAS NO SUBCONTINUUM*

BY

R. L. WILDER

It is well known† that every closed, connected and regular (connected im kleinen)‡ point set is arc-wise§ connected. That a connected and regular point set which is not closed is not necessarily arc-wise connected, and, indeed, may contain no arc, was shown by R. L. Moore.‖ Since an arc is a very special kind of continuum,¶ Moore's result suggests the possibility of the existence of connected and regular point sets which not only fail to contain any arc, but do not contain any continuum whatsoever. It is the purpose of this paper to establish the existence of such sets.

I shall make use of an example of a connected set S containing a point a such that $S-a$ is totally disconnected, given by B. Knaster and C. Kuratowski.** For the sake of brevity, I shall assume familiarity with this set as well as with the proof of its connectivity.

I

Before proceeding to the construction of the point set which I wish to present in this paper, I shall first construct, and establish certain properties

---

* Presented to the Society, April 2, 1926; received by the editors in May, 1926.

† Cf. R. L. Moore, *A theorem concerning continuous curves*, Bulletin of the American Mathematical Society, vol. 23 (1917), pp. 233–236; S. Mazurkiewicz, *Sur les lignes de Jordan*, Fundamenta Mathematicae, vol. 1 (1920), pp. 166–209; H. Tietze, *Über stetige Kurven, Jordansche Kurvenbogen und geschlossene Jordansche Kurven*, Mathematische Zeitschrift, vol. 5 (1919), pp. 284–291.

‡ A point set M is said to be *regular* or *connected im kleinen at a point* P if for every circle K with center at P there exists a circle R concentric with K such that if x is a point of M interior to R, then P and x both lie in a connected subset of M which lies wholly interior to K. The set M is said to be *regular* or *connected im kleinen* if it is regular at every one of its points. Cf. H. Hahn, *Über die allgemeinste ebene Punktmenge, die stetiges Bild einer Strecke ist*, Jahresbericht der Deutschen Mathematiker-Vereinigung, vol. 23 (1924), pp. 318–322. Also S. Mazurkiewicz, loc. cit.

§ A point set M is said to be *arc-wise connected* if every two of its points are the end points of at least one arc of M.

‖ See Bulletin of the American Mathematical Society, vol. 29 (1923), p. 438.

¶ A *continuum* is a closed and connected point set which contains more than one point.

** *Sur les ensembles connexes*, Fundamenta Mathematicae, vol. 2 (1921), pp. 206–255. The set S referred to here is described on pp. 241–244. Hereafter, I shall refer to this paper as S.E.C.

of, a set of which a certain important subset is in one-to-one continuous correspondence with the set $S-a$.

On the interval $[0, 1]$ of the x-axis, let X be a non-dense perfect set. On all lines $x=\xi$, where ξ is the abscissa of an end point of an interval complementary to X, consider those points whose ordinates are rational numbers y such that $0 \leqq y < 1$. Call the set of all such points P_1. Let P_2 be the set consisting of all points whose ordinates are irrational numbers between 0 and 1 on lines $x=\xi$, where ξ is the abscissa of a point of X which is not an end point of an interval complementary to X.

There exists a one-to-one continuous correspondence between the set P_1+P_2 and the set $S-a$, which may be defined in the following manner: Let the set X described above and the set C as defined in §5 of S.E.C. be identical. For each point P of $S-a$ let x_P denote the abscissa of the point in which the x-axis is intersected by the line from P to a and let y_P denote twice the ordinate of P. There exists a correspondence between $S-a$ and P_1+P_2 in which to each point P of $S-a$ there corresponds the point of P_1+P_2 whose abscissa and ordinate are x_P and y_P respectively. This correspondence is continuous and will hereafter be called the correspondence Z.

As the segments* complementary to X form a denumerable set, they can be ordered in a sequence $e_1, e_2, e_3, \cdots$. The set of all points which lie above e_n $(n=1, 2, 3, \cdots)$ and below $y=1$ denote by s_n. Let $Q_j, j=2^{n-1}, 2^{n-1}+1, \cdots, 2^n-1$, be 2^{n-1} connected sets such that if L_i is the line $y=i/2^n$, where i takes on all odd integral values between 0 and 2^n, and P_{1in} and P_{2in} are the end points of that segment of L_i whose projection on the x-axis is e_n, then for $j=2^{n-1}+\frac{1}{2}(i-1)$, (1) Q_j contains P_{1in} and P_{2in}, is regular at these points, and, except for these two points, lies wholly in s_n; (2) all the limit points of Q_j except P_{1in} and P_{2in} lie in s_n; (3) if x is any point of Q_j, the distance of x from L_i is less than $1/2^n$.†

Let

$$M = P_1 + P_2 + \sum_{i=1}^{\infty} Q_i.$$

DEFINITION. If A and B are two distinct points of a point set N, then A and B are *separated in N in the weak sense* provided N contains no connected subset which contains both A and B. The points A and B are *sepa-*

* By segment I mean an interval without its end points.

† This number is purely conventional, since all that is necessary for the proof below is that the diameter of Q_j approach zero as j increases.

rated in N in the strong sense if there exist two mutually separated* sets whose sum is N and which contain A and B, respectively.

The following lemma is part of a theorem which I have established in another connection, but which is not yet published:

LEMMA 1. *If N is a continuous curve,† and A and B are two distinct points of an open subset K of N which are not separated in K in the strong sense, then A and B are not separated in K in the weak sense.*

LEMMA 2. *Let J be a plane simple closed curve which is the sum of two arcs AxB and AyB which have only their end points A and B in common, R the bounded domain complementary to J, and C a continuum which separates the plane between A and B.‡ Then there exists a continuum C' which is a subset of C and of $J+R$, and which has at least one point in common with each of the arcs AxB and AyB.*

Suppose no such continuum as C' exists. For every point P of C in R let $M(P)$ denote that maximal connected subset§ of $C\times(J+R)$ determined by P. Every set $M(P)$ has at least one point in common with one of the arcs AxB and AyB‖ but not with both. Every point P whose corresponding $M(P)$ contains a point of AxB assign to a set M_1, and every point P whose corresponding $M(P)$ contains a point of AyB assign to a set M_2. There exists on AxB an arc $A'xB'$ which does not contain A or B, but which does contain every point of the set common to C and AxB. Similarly there exists on AyB an arc $A''yB''$ which does not contain A or B, but which contains every point common to C and AyB.

Let

$$A'xB' + M_1 = N_1,$$
$$A''yB'' + M_2 = N_2.$$

N_1 and N_2 are continua neither of which separates¶ A from B in $J+R$.

---

* Two point sets are said to be *mutually separated* if they are mutually exclusive and neither contains a limit point of the other.

† A *continuous curve* is a bounded, closed, connected and regular point set. A subset K of a continuous curve N is an *open subset* of N provided $N-K$ is a closed, non-vacuous set.

‡ I.e., there exists no connected subset in the plane which contains A and B and no point of C.

§ If P is a point of a point set M, then that maximal connected subset of M determined by P is the set of all points of M that lie with P in a connected subset of M.

‖ Cf. Anna M. Mullikin, *Certain theorems relating to plane connected point sets*, these Transactions, vol. 24 (1922), pp. 144–162.

¶ A subset N of a point set M is said to *separate* two points A and B of $M-N$, if $M-N$ is the sum of two mutually exclusive sets T and U which contain A and B respectively, and neither of which contains a limit point of the other.

Hence their sum, N_1+N_2, does not separate A from B in $J+R$.* That is, A and B are two points of the open subset $J+R-(N_1+N_2)$ of the continuous curve $J+R$, which are not separated in that open subset in the strong sense. Hence by Lemma 1, A and B are not separated in that open subset in the weak sense. Accordingly, there exists a connected subset of $J+R-(N_1+N_2)$ which contains A and B. As N_1+N_2 contains all points common to $J+R$ and C, a contradiction of the fact that C separates the plane between A and B results.

LEMMA 3. *The point set M described above is connected.*

Suppose M is not connected. Then it is the sum of two sets M_1 and M_2 which are mutually separated. Clearly no Q_i has points in both M_1 and M_2, as every Q_i is connected. Hence each of the sets M_1, M_2 contains points of the set P_1+P_2.

Let

$$M_1 \times (P_1 + P_2) = T_1,$$

$$M_2 \times (P_1 + P_2) = T_2.$$

T_1 and T_2 are mutually separated sets, since they are subsets of mutually separated sets, and

$$T_1 + T_2 = P_1 + P_2.$$

If m is a point of P_1+P_2 in T_1, say, then all those points of P_1+P_2 on the line $x=\xi_1$ which contains m belong to T_1. For suppose that q is a point of P_1+P_2 on $x = \xi_1$ belonging to T_2. Let the ordinate of q be less than that of m. By Theorem 37 (p. 233) of S.E.C., there exists a continuum C which contains no point of P_1+P_2 and which separates the plane between m and q.

Denote the intersection of $x=\xi_1$ and the x-axis by z. Then z is a limit point from the left, say, of end points of complementary intervals of X. Hence there exists an infinite sequence of distinct segments of the set $[e_n]$, viz., $\bar{e}_1, \bar{e}_2, \bar{e}_3, \cdots$, whose end points have z as a sequential limit point,† and all of which lie to the left of z. Denote the corresponding elements of

---

* Cf. R. L. Moore, *Concerning upper semi-continuous collections of continua which do not separate a given continuum*, Proceedings of the National Academy of Sciences, vol. 10 (1924), pp. 356–360, Theorem 2.

† A point P is a sequential limit point of a point set K if every circle which encloses P also encloses all except a finite number of points of K.

the set $[s_n]$ by $\bar{s}_1, s_2, \bar{s}_3, \cdots$. Let b be a point whose abscissa is less than that of m, and whose ordinate is the same as that of m. Let d be a point which has the same abscissa as b and the same ordinate as q. Denote the interior of the rectangle $bmqdb$ by R. By Lemma 2, there exists a subcontinuum C' of C which contains points of the broken line $mbdq$ and of the straight line interval mq, and which, except for these points, is a subset of R.

Since m and q are not on C', and C' is closed, there exists a number $\eta<\xi_1$, such that if r and u are the intersections of $x=\eta$ with the straight line intervals bm and dq, respectively, there are no points of C' on the straight line intervals rm, uq. Also, there exist two elements $\bar{s}_i$ and $\bar{s}_j$ of the set $[\bar{s}_n]$ such that the end points of $\bar{e}_i$ and $\bar{e}_j$ lie between the point on the x-axis whose abscissa is η and the point z. Let (f) and (g) be abscissas of points in $\bar{e}_i$ and $\bar{e}_j$, respectively, and suppose $(f)<(g)$. There exists a subcontinuum $\overline{C}$ of C' which contains points of the lines $x=(f)$ and $x=(g)$, and which, except for these points, lies wholly between these two lines. As $\eta<(f)<(g)<\xi_1$, $\overline{C}$ lies wholly in R. Let the highest point of $\overline{C}$ on $x=(f)$ be f, and on $x=(g)$ be g, and let the projections of these two points on the x-axis be F and G, respectively. Let the smallest ordinate possessed by any point on $\overline{C}$ be ϵ, and let the intersections of $y=\epsilon$ with $x=(f)$ and $x=(g)$ be $\bar{f}$ and $\bar{g}$ respectively. Let those points of P_1+P_2 which lie within the rectangle $f\bar{g}GF\bar{f}$ be denoted by $\overline{P}_1$. The set $\overline{P}_1$ is certainly not vacuous, for if ρ is the abscissa of the right-hand end point of $\bar{e}_i$, there are points of $\overline{P}_1$ on $x=\rho$. Let $\overline{P}_2$ denote the set of points of P_1+P_2 which lie between $x=(f)$ and $x=(g)$, and above $\overline{C}$.

Let w be any arc whose end points are F and G and which lies, except for F and G, entirely below the x-axis. Then the arc w, together with the straight line intervals fF and gG and the continuum $\overline{C}$, forms a continuum K which separates the plane between $\overline{P}_1$ and $\overline{P}_2$ and contains no point of P_1+P_2. Let H denote the set of all those points of P_1+P_2 which lie in that complementary domain of K which contains $\overline{P}_1$. The sets H and P_1+P_2-H are mutually separated. Let H^* and $(P_1+P_2-H)^*$ denote the subsets of $S-a$ which correspond to H and P_1+P_2-H respectively under the correspondence Z described above. It is easy to see that a is not a limit point of H. Hence $a+(P_1+P_2-H)^*$ and H^* are mutually separated. This is impossible, since S is connected. Thus the supposition that q is a point of T_2 leads to a contradiction and all points of P_1+P_2 on a line $x=\xi$ belong wholly to T_1, or to T_2, as the case may be.

Let the set of those points of X which are the projections of points of

$T_i (i=1, 2)$ on the x-axis be denoted by X_i. Then X_1 and X_2 are mutually separated.

If one end point of a segment e_n is in X_1, say, both end points are in X_1. Otherwise the point P_{11n} would be in T_1, say, and P_{21n} would be in T_2, which is impossible since Q_{2n-1} lies wholly in one of the sets M_1, M_2. For every n, add all points of e_n to that set X_i to which its end points belong. Denote the sets obtained by adding points to X_1 and X_2 in this way by U_1 and U_2, respectively. Then U_1 and U_2 are mutually separated, since no point of a segment e_n is a limit point of points of the x-axis exterior to e_n. But this is impossible, since the set U_1+U_2 is identical with the set of all points in the interval $[0, 1]$. Hence the supposition that M is not connected leads to a contradiction.

LEMMA 4. *The point set M is regular at every point of the set P_1+P_2.*

Let t be a point of P_1+P_2 which does not belong to any Q_i. Let K_1 be any circle with center at t. Let W be a rectangle with center at t, lying wholly interior to K_1, with sides parallel to the coördinate axes and such that (1) the vertical sides of W contain no points of P_1+P_2, (2) if H is the height of W, the width of W is such that no elements of the set $[s_n]$ have points within W whose subscripts, n, do not satisfy the inequality $1/2^n<H/4$, it being understood that this condition does not apply to the elements of $[s_n]$ in which the vertical sides lie, (3) the upper and lower bases of W lie at a rational distance above the x-axis. Let K_2 be a circle concentric with K_1, lying interior to W, and not enclosing any points of those elements of the set $[Q_n]$ which have points on W. Such a circle exists, since t is not a point or a limit point of any Q_i and the diameters of the sets $[Q_n]$ approach zero as n increases. Clearly no point of P_1+P_2 interior to W is a limit point of the sum of all those sets of $[Q_n]$ which have points on W, unless it be a point of some such set.

Let N be the set of all points of P_1+P_2 interior to W or on the lower base of W, together with all sets Q_n which lie wholly interior to W. Clearly N is similar to the set M and it may be shown to be connected by the same argument that was used to prove M connected. As K_2 encloses only those points of M which belong to N, M is obviously regular at t.

If t is a point of P_1+P_2 belonging to a set Q_j, select W as before. Select K_2 so that it encloses no points of any set Q_n excepting Q_j, which has points on W. Select N as before. Since Q_j is regular at t, there exists a circle K_3, concentric with K_1, such that all points of Q_j interior to K_3 lie, with t, in a connected subset of Q_j which lies wholly interior to K_1. Then the smaller

of the two circles K_2, K_3(or either of them, if they are identical) is clearly such that any point of M within it lies, with t, in a connected subset of M which lies wholly interior to K_1.

II

Consider a set N constructed as follows: On the interval $[0, 1]$ of the x-axis, let X be a non-dense perfect set. Denote the point $(\frac{1}{2}, 1)$ by P. For any point x of X, if x is an end point of an interval complementary to X, assign the straight line interval Px to a class E_1; otherwise to a class E_2. The set of all points on the xy-plane whose ordinates are rational and which lie on intervals of E_1 form a point set N_1, and the set of all points whose ordinates are irrational and which lie on intervals of E_2 form a set N_2. The set N_1+N_2 is connected, as it is identical, except for choice of coördinates, with the set S. Let the set N_1+N_2, together with the set of points symmetrical to it with respect to the x-axis, be denoted by Q_0. The set Q_0 is obviously connected. Let the point symmetrical to P with respect to the x-axis be denoted by P'.

As the intervals complementary to X on the interval $[0, 1]$ form a denumerable set, they can be ordered in a sequence $I_1, I_2, I_3, \cdots$. For every n, let e_{1n} and e_{2n} denote the end points of I_n, and let the set of points interior to the triangle $Pe_{1n}e_{2n}$ be denoted by s_n. Let Q_k^1, $k=2^{n-1}, 2^{n-1}+1, \cdots, 2^n-1$, be 2^{n-1} sets such that (1) for every k, Q_k^1 is in one-to-one continuous correspondence with Q_0, (2) if L_i is the line $y=i/2^n$, where i takes on all odd integral values between 0 and 2^n, and P_{1in} and P_{2in} are the intersections of $y=i/2^n$ with the straight line intervals Pe_{1n} and Pe_{2n}, respectively, then for $k=2^{n-1}+\frac{1}{2}(i-1)$, Q_k^1 (i) contains P_{1in} and P_{2in} and these points correspond, in the one-to-one continuous correspondence between this set and Q_0, to P and P', respectively, and (ii) is symmetrical with respect to the line $y=i/2^n$, has no limit points, except P_{1in} and P_{2in}, that do not lie in s_n, and if x is any point of it, the distance of x from L_i is less than $1/2^n$.

Denote the class each of whose elements is a set Q_i^1 $(i=1, 2, \cdots)$ by $[Q^1]^1$ and by $[Q^1]^2$ the class whose elements are symmetrical, with respect to the x-axis, to the elements of $[Q^1]^1$, and by $[Q^1]$ the class whose elements are the elements of $[Q^1]^1$ and $[Q^1]^2$.

Denote by $[Q^2]$ a class whose elements are constructed relative to *all* the elements of $[Q^1]$ just as the elements of $[Q^1]$ were constructed relative to Q_0. In general, denote by $[Q^n]$ a class whose elements are constructed relative to all the elements of $[Q^{n-1}]$ just as the elements of $[Q^1]$ were constructed relative to Q_0.

Let N be that point set which consists of all points of Q_0, as well as of all points which are contained in elements of the classes $[Q^i]$ $(i=1, 2, 3, \cdots)$.

THEOREM. *The point set N is a connected regular point set which contains no continuum.*

That N is connected is evident. That N is regular at P and P' is easily shown from the fact that Q_0 is itself regular at these points. Similarly, every element of $[Q^n]$ $(n=1, 2, 3, \cdots)$ is regular at the points corresponding to P and P'. The set $N-(P+P')$ is therefore a set of the same type as the set M described above, and is therefore regular at all points of the set $Q_0-(P+P')$ by Lemma 4.

For every n, arrange the elements of the class $[Q^n]$ in a sequence Q_1^n, Q_2^n, Q_3^n, $\cdots$. If Q_i^1 is any element of $[Q^1]$, it has been shown already that N is regular at the points of Q_i^1 corresponding to P and P'. As for the other points of Q_i^1, none of these is a limit point of the other elements of $[Q^1]$, and as Q_i^1 is a set of the same type as the set Q_0, N is regular at all these points. As for any element Q_i^n of any class $[Q^n]$, a similar argument holds, since no one of these sets contains, except for the points corresponding to P and P', a limit point of any element or set of elements of the classes $[Q^i]$ $(i=1, 2, 3, \cdots, n-1)$ or of $[Q^n]$ distinct from itself. Hence N is a regular point set.

The point set N contains no continuum. For suppose it does contain a continuum K. Then K contains a point, A, of an element Q_i^j of the class $[Q^j]$, say, where A is not one of the two points of Q_i^j corresponding to P and P' of Q_0. For the set of points in sets Q_i^j $(i=1, 2, \cdots, j=1, 2, \cdots)$ corresponding to P and P' of Q_0 is clearly denumerable and no continuum, nor indeed any connected set, can consist of a denumerable set of points. If d and f are the points of Q_i^j corresponding to P and P' in Q_0, it is easily shown by Theorem 1 of the paper of Miss Mullikin referred to above that K has a subcontinuum, K_0, which contains A, but does not contain d or f, nor any point of any other element of the class $[Q^j]$, nor any point of an element of a class $[Q^n]$ for which $n<j$. The continuum K_0 cannot lie wholly in Q_i^j since the latter set contains no continuum. Hence K_0 contains a point B of some set Q_k^m such that B is not a point of this set corresponding to P or P' in Q_0, and such that $m>j$. Then K_0 contains a subcontinuum K_1 which contains B but no point of an element of a class $[Q^n]$ for which $n<m$. By repetition of this argument, the existence of a sequence of continua

$K_1, K_2, K_3, \cdots$ may be shown such that for every positive integer q, K_{q+1} is a subset of K_q, and such that K_q contains no point of any element of a class $[Q^n]$ for which $n<q$. The sets $K_1, K_2, K_3, \cdots$ have in common at least one point P, and this point does not belong to any element of a set $[Q^n]$. Thus the supposition that N contains a continuum K is shown to lead to a contradiction, and the theorem is proved.

OHIO STATE UNIVERSITY,
COLUMBUS, OHIO

MEROMORPHIC FUNCTIONS WITH ADDITION OR MULTIPLICATION THEOREMS*

BY

J. F. RITT

It is a theorem of Briot and Bouquet that if a solution of an algebraic differential equation of the first order,

$$F\left(\frac{dw}{dz},\ w\right) = 0,$$

is uniform, the solution is a rational function, or a rational function of $e^{\mu z}$ (μ constant), or an elliptic function. It is part of a famous theorem of Weierstrass that every uniform function with an algebraic addition theorem is of one of the three types just enumerated. It is a result of Picard that if two meromorphic functions are algebraically related to one another, either the relation is of genus zero, in which case the two functions are rational functions of a single meromorphic function, or the relation is of genus unity, in which case the two functions are elliptic functions of a single integral function.†

The present paper treats a problem which might easily be suggested by any of the above theorems. We seek all meromorphic functions $f(z)$ such that, for some linear $\lambda(z)$, an algebraic relation

$$F(f[\lambda(z)],\ f(z)) = 0$$

holds.

There is no generality lost in assuming $\lambda(z)$ to be of one of the forms $z+m$ or mz. Thus our problem is related to difference and q-difference equations as that of Briot and Bouquet is to differential equations. Of course, the assumption made here that $f(z)$ is meromorphic is stronger than the requirement of mere uniformity in the differential equation problem. For the case in which $\lambda(z)=z+m$, we have, as we are dealing with a single value

---

* Presented to the Society, February 27, 1926; received by the editors in March, 1926.

† Acta Mathematica, vol. 11 (1887), p. 1. Picard's theorem deals with any pair of uniform functions with an isolated essential singularity. See also Bulletin des Sciences Mathématiques, vol. 15 (1880), p. 416, where an incomplete proof of the theorem is given for meromorphic functions. The proof in that paper that when the relation is of genus unity the functions are elliptic functions of an integral function, is rigorous.

of m, a generalization of Weierstrass's addition theorem problem, to be sure with a heavy function-theoretic restriction. Of course, our investigation also amounts to the further study of the two functions in Picard's theorem when one of the functions is found from the other by transforming the variable linearly.

In stating our results, we assume that $\lambda(z)=mz$ with m a constant not 0 or 1, or that $\lambda(z)=z+m$ with m a constant not 0. As has already been said, this involves no sacrifice of generality.

The relation between $f(z)$ and $f[\lambda(z)]$ is, by Picard's theorem, either of genus unity or of genus zero. For the case in which $f(z)$ and $f[\lambda(z)]$ satisfy a relation of genus unity, we prove that $f(z)$ is an elliptic function of an integral function $g(z)$, *where* $g[\lambda(z)]$ *is a linear function of* $g(z)$. Explicit expressions are given for $g(z)$ in §4.

The case of genus zero is more complicated. There are the following possibilities:

(a) $f(z)$ *is an elliptic function of an integral function* $g(z)$, *with* $g[\lambda(z)]$ *a linear function of* $g(z)$.

(b) $f(z)$ *is a rational function of a meromorphic function* $h(z)$, *where some one of the functions* $h(z)$, $h[\lambda(z)]$ *is a rational function of the other.*

If $\lambda(z)=mz$, $h(z)$ belongs to a class of functions discussed by Poincaré in a paper on rational multiplication theorems.* Poincaré's work tells enough about these functions for them to be regarded as known. If $\lambda(z)=z+m$, $h(z)$ is a function with a *rational addition theorem*. No complete determination of such functions has thus far been made. A large class of them can be obtained from Poincaré's functions by a substitution, but it is by no means certain that they are all found in this manner.

(c) $f(z)$ *is a rational function of* $h[(g(z))^{1/n}]$, *where* n *is an integer,* $g(z)$ *an integral function such that* $(g[\lambda(z)])^{1/n}$ *is linear in* $(g(z))^{1/n}$, *and where* $h(z)$ *is one of the following periodic meromorphic functions:* e^z, $\cos z$, $\wp(z)$; *in the lemniscatic case* $(g_3=0)$, $\wp^2(z+u)$; *in the equianharmonic case* $(g_2=0)$, $\wp'(z)$ *and* $\wp^3(z+u)$.

There are in each case a finite number of possibilities for n and u, which are stated in detail in II. The possibilities for $g(z)$ are also determined. Both $h[(g(z))^{1/n}]$ and $h[(g[\lambda(z)])^{1/n}]$ will be rational in a third function $h[p(g(z))^{1/n}$

* Journal de Mathématiques, (3), vol. 55 (1890).

$+q]$, but neither of the first two functions will be a rational function of the other.*

Case (c) provides most of the work in this paper. It turns out that our problem is in a close relation with the problem of determining all pairs of rational functions $\Phi(z)$ and $\Psi(z)$ such that $\Phi[\Psi(z)]=\Psi[\Phi(z)]$, a question which the writer has studied in an earlier paper.† Two certain sequences of rational functions which occur in the permutable function problem appear again in this paper. Direct application is made here of the information previously obtained in regard to those sequences.

A question which we do not settle is whether Case (a) gives any functions $f(z)$ not given by Cases (b) and (c). This depends on whether certain algebraic relations derived from the relation between $f(z)$ and $f[\lambda(z)]$ are of genus unity rather than of genus zero.‡ Except for this point, and for the question raised under Case (b) as to functions with rational addition theorems, our investigation is complete.

I. Relations of genus unity

1. **Discussion of $\lambda(z)$.** Let $f(z)$ be a non-rational meromorphic function and $\lambda(z)$ a non-constant and non-identical linear function. If $f(z)$ and $f[\lambda(z)]$ satisfy an algebraic relation, then $\lambda(z)$ must be integral, that is $\lambda(z)=mz+n$ with m and n constant. For if $\lambda(z)$ had a pole at a finite point α, $f[\lambda(z)]$, and therefore also $f(z)$, would have an essential singularity at α.

If $m\neq 1$, we write

$$f_1(z) = f\left(z + \frac{n}{1-m}\right).$$

Then $f_1(z)$ and $f_1(mz)$ obey the same relation as do $f(z)$ and $f[\lambda(z)]$. We therefore assume that $\lambda(z)$ is of one of the forms mz or $z+m$, throwing away no generality in the assumption.

2. **Relations of genus unity.** We assume in this section that $f(z)$ and $f[\lambda(z)]$ satisfy a relation of genus one.

Let $\lambda_2(z)=\lambda[\lambda(z)]$. Evidently $f[\lambda_2(z)]$ is algebraic in $f(z)$.

---

* For instance, the relation between $\cos(4z)^{1/2}$ and $\cos z^{1/2}$ comes under Case (b), whereas that between $e^{2z/2}$ and e^z comes under Case (c).

† Ritt, *Permutable rational functions*, these Transactions, vol. 25 (1923), p. 399.

‡ See §6.

As $f[\lambda(z)]$ and $f[\lambda_2(z)]$ are both algebraic in $f(z)$, we can find a rational combination of $f[\lambda(z)]$ and $f[\lambda_2(z)]$—call it $h(z)$—such that $f[\lambda(z)]$ and $f[\lambda_2(z)]$ are both rational combinations of $f(z)$ and $h(z)$.*

Then $f(z)$ and $h(z)$ are meromorphic functions between which there exists an algebraic relation. This relation cannot be of genus zero. If it were, $f(z)$ and $h(z)$ would both be rational functions of a single meromorphic function. Consequently $f[\lambda(z)]$, rational in $f(z)$ and $h(z)$, would also be rational in this new meromorphic function, so that the relation between $f(z)$ and $f[\lambda(z)]$ would have to be of genus zero.

Hence $f(z)$ and $h(z)$ satisfy a relation of genus unity, and are elliptic functions of an integral function $g(z)$. Then $f[\lambda(z)]$ and $f[\lambda_2(z)]$ are also elliptic functions of $g(z)$. Let

$$(1) \qquad f(z) = \phi_1[g(z)], \quad f[\lambda(z)] = \phi_2[g(z)], \quad f[\lambda_2(z)] = \phi_3[g(z)],$$

where $\phi_1(z)$, $\phi_2(z)$, and $\phi_3(z)$ are elliptic functions.

Let the relation of genus unity which connects $f(z)$ and $f[\lambda(z)]$ be

$$(2) \qquad F(u, v) = 0,$$

that is, let (2) be satisfied when u and v are replaced by $f(z)$ and $f[\lambda(z)]$ respectively. Then (2) is satisfied when u is replaced by $\phi_1(z)$ and v by $\phi_2(z)$. Since $f[\lambda(z)]$ and $f[\lambda_2(z)]$ obey (2), we can also satisfy (2) by replacing u and v by $\phi_2(z)$ and $\phi_3(z)$ respectively.

Let

$$\int R(u, v)du$$

be an integral of the first kind attached to the relation (2). If, in the expression

$$R(u, v)\frac{du}{dz},$$

we replace u and v by $\phi_1(z)$ and $\phi_2(z)$ respectively, or by $\phi_2(z)$ and $\phi_3(z)$ respectively, we obtain an elliptic function whose integral with respect to z could not become infinite for any finite z.† The elliptic function thus obtained must be a constant.

---

* Weber, *Algebra*, 2d edition, vol. 1, p. 499.

† Cf. Picard, *Traité d'Analyse*, 2d edition, vol. 2, p. 561.

We shall replace u and v in (2) by $f[\lambda(z)]$ and $f[\lambda_2(z)]$ respectively. Using (1), we see that the result is a constant times the derivative of $g(z)$. But if we use the expressions

$$f[\lambda(z)] = \phi_1(g[\lambda(z)]), \qquad f[\lambda_2(z)] = \phi_2(g[\lambda(z)]),$$

we find the same result expressed as a constant times the derivative of $g[\lambda(z)]$.

Hence

$$\frac{d}{dz}\, g[\lambda(z)] = p\, \frac{d}{dz} g(z)$$

with p a constant, so that

$$g[\lambda(z)] = pg(z) + q, \tag{3}$$

with q constant, as well as p. We have thus, for the case of genus unity, the result stated in the introduction.

3. **The constants p and q.** Equation (3) leads quickly to explicit expressions for $g(z)$. Before giving these expressions, let us examine the constants p and q, assuming $\phi_1(z)$ to have been given as any elliptic function.

We shall fix p and q so that $\phi_1(z)$ and $\phi_1(pz+q)$ may be algebraically related, without insisting that the relation between them be of genus unity. It is necessary and sufficient that $\phi_1(z)$ and $\phi_1(pz+q)$ have a pair of periods in common. There is no restriction on q. Any rational value of p is admissible, and when the ratio of the periods of $\phi_1(z)$ is a quadratic number, there is a system of complex values for p.

If, $\lambda(z)$ being given, we use every elliptic function $\phi_1(z)$, every p and q described as above, and every integral $g(z)$ satisfying (3), we shall obtain every meromorphic $f(z)$ which satisfies with $f[\lambda(z)]$ a relation of genus unity. But for certain choices of $\phi_1(z)$, p and q, the relation will be of genus zero.*

4. **The function $g(z)$.** We shall now discuss $g(z)$ and $\lambda(z)$, regarding p and q as known. The results will be used again in II.

Suppose first that $\lambda(z) = mz$. Let $g(z)$ have, for $z=0$, the development

$$g(z) = a + bz^n + cz^{n+1} + \cdots,$$

where $b \neq 0$.

We must have $m^n = p$. Suppose first that p (and therefore m) is not a rational root of unity. All terms after the second in $g(z)$ disappear, and we have

---

* For instance, let $\phi_1(z)$ be $\wp(z)$, p be rational and q be zero.

$$g(z) = \frac{q}{1-p} + bz^n,$$

with b and n arbitrary.

Let p be a root of unity. Taking any positive integral n, we may choose for m any nth root of p. Let m thus chosen be a primitive ith root of unity. We must have

$$g(z) = a + z^n h(z^i),$$

with $h(z)$ an integral function. We cannot have $p=1$, unless $q=0$. When $p\neq 1$, we have $a=q/(1-p)$, whereas n and $h(z)$ are arbitrary. If $p=1$, $q=0$, a becomes arbitrary.

Suppose now that $\lambda(z)=z+m$. When $p\neq 1$, we find, by elementary considerations, that

$$g(z) = e^{z(\log p)/m}\pi(z) - \frac{q}{p-1},$$

where $\pi(z)$ is any periodic integral function with m for period. Any value distinct from zero can be used for m. When $p=1$, we find that

$$g(z) = \frac{q}{m}z + \pi(z)$$

with $\pi(z)$ and m as above.

II. Relations of genus zero

5. **The functions $f_i(z)$.** If $\phi(z)=\phi_1[\phi_2(z)]$, where $\phi_1(z)$ and $\phi_2(z)$ are rational and $\phi_2(z)$ is not linear, we shall call $\phi_2(z)$ a *forefactor* of $\phi(z)$.

Let the meromorphic functions $f_0(z)$ and $f_0[\lambda(z)]$, λ linear, satisfy a relation of genus zero. We have

$$f_0(z) = \phi_0[f_1(z)], \qquad f_0[\lambda(z)] = \psi_0[f_1(z)],$$

with $f_1(z)$ meromorphic and $\phi_0(z)$ and $\psi_0(z)$ rational.

It is possible so to choose $f_1(z)$ that $f_1(z)$ is a rational function of $f_0(z)$ and $f_0[\lambda(z)]$.* We suppose this to be done. Then $\phi_0(z)$ and $\psi_0(z)$ have no forefactor in common. For if they had a common forefactor $\sigma(z)$, then $f_1(z)$, as a rational combination of $f_0(z)$ and $f_0[\lambda(z)]$, would be a rational function of $\sigma[f_1(z)]$.

As

$$f_0[\lambda(z)] = \psi_0[f_1(z)] = \phi_0(f_1[\lambda(z)]),$$

* Appell et Goursat, *Théorie des Fonctions Algébriques*, Paris, 1895, p. 288.

$f_1(z)$ and $f_1[\lambda(z)]$ are algebraically related. The relation which they satisfy is of genus unity or of genus zero. If it is of genus unity, $f_1(z)$ is an elliptic function of an integral function $g(z)$, where $g(z)$ is described as in §4. Of course, $f_0(z)$ will also be an elliptic function of $g(z)$.

Suppose that $f_1(z)$ and $f_1[\lambda(z)]$ satisfy a relation of genus zero. Then there exist a meromorphic function $f_2(z)$, and two rational functions, $\phi_1(z)$ and $\psi_1(z)$, with no forefactor in common, such that

$$f_1(z) = \phi_1[f_2(z)], \qquad f_1[\lambda(z)] = \psi_1[f_2(z)].$$

6. **Case** (a). We continue the above procedure. It is conceivable that, for certain functions $f_0(z)$, we reach at some step an $f_i(z)$ which satisfies with $f_i[\lambda(z)]$ a relation of genus unity. If such functions $f_0(z)$ exist, they are elliptic functions of integral functions.

Whether a relation of genus unity is ever reached for certain functions, is a question which we do not settle in this paper. But certainly there exist functions $f_0(z)$ which are elliptic functions of integral functions and which satisfy with $f_0[\lambda(z)]$, for some non-identical $\lambda(z)$, a relation of genus zero. This makes it legitimate to create Case (a) under a genus zero, in our statement of results in the introduction. With Case (a) established, our results assume a simple form. But if it should be that no relation of genus zero leads to a relation of genus unity, then Case (a) may be suppressed.

7. **The sequences** (I) **and** (II). We shall assume in what follows that at no step are we led to a relation of genus unity. Then there exist two infinite sequences of rational functions

$$\text{(I)} \qquad \phi_0(z), \quad \phi_1(z), \quad \phi_2(z), \cdots, \quad \phi_i(z), \cdots,$$

$$\text{(II)} \qquad \psi_0(z), \quad \psi_1(z), \quad \psi_2(z), \cdots, \quad \psi_i(z), \cdots,$$

where, for every i, $\phi_i(z)$ and $\psi_i(z)$ have no common forefactor, and an infinite sequence of meromorphic functions

$$f_0(z), \quad f_1(z), \quad f_2(z), \cdots, \quad f_i(z), \cdots,$$

such that, for every i,

$$f_i(z) = \phi_i[f_{i+1}(z)], \quad f_i[\lambda(z)] = \psi_i[f_{i+1}(z)].$$

As

$$f_i[\lambda(z)] = \phi_i(f_{i+1}[\lambda(z)]) = \phi_i(\psi_{i+1}[f_{i+2}(z)]),$$
$$f_i[\lambda(z)] = \psi_i[f_{i+1}(z)] = \psi_i(\phi_{i+1}[f_{i+2}(z)]),$$

we have, for every i,

$$\phi_i[\psi_{i+1}(z)] = \psi_i[\phi_{i+1}(z)].$$

8. **The degrees r and s.** At this point we shall begin to use the methods and results of our paper *Permutable rational functions*, to which reference has already been made. This paper will be called A.

We represent the degrees of $\phi_i(z)$ and $\psi_i(z)$ by r_i and s_i respectively. We shall prove that, for every i,

$$r_{i+1} \leqq r_i, \qquad s_{i+1} \leqq s_i.$$

We write

$$w = G(z) = \phi_i[\psi_{i+1}(z)] = \psi_i[\phi_{i+1}(z)].$$

The function $\psi_{i+1}(z)$ determines r_i systems of imprimitivity of the group of monodromy of $G^{-1}(w)$, each consisting of s_{i+1} letters.* Similarly $\phi_{i+1}(z)$ determines s_i such systems, each system consisting of r_{i+1} letters. If we had $r_{i+1} > r_i$, some system determined by $\phi_{i+1}(z)$ would have more than one letter in common with at least one of the systems determined by $\psi_{i+1}(z)$. This would mean that $\phi_{i+1}(z)$ and $\psi_{i+1}(z)$ have a common forefactor.† Hence $r_{i+1} \leqq r_i$, $s_{i+1} \leqq s_i$.

9. **Paper A.** There occur in A, two sequences of rational functions, called (A) and (B),

$$(A) \qquad \phi_0, \ \phi_1, \ \phi_2, \ \cdots, \qquad (B) \qquad \psi_0, \ \psi_1, \ \psi_2, \ \cdots,$$

whose theory is practically identical with that of the sequences (I) and (II) of this paper. In studying the sequences (A) and (B), we made use of the following facts:

(a) *The degrees of the functions in both sequences are non-increasing.*

(b) *For every i, ϕ_i and ψ_i have no common forefactor.*

(c) *For every i, $\phi_i\psi_{i+1} = \psi_i\phi_{i+1}$.*‡

From (a) it followed immediately that, for sufficiently large values of i, the degrees of ϕ_i and ψ_i do not change with i. We proved also, using (b) and (c), that there exist two integers, $h \leqq 4$ and $k \leqq 4$, such that, for i sufficiently large, the inverse of ϕ_i has precisely h critical points and that of ψ_i precisely k critical points. We then determined, with the help of (b) and (c), the nature of the critical points of the inverses ϕ_i^{-1} and ψ_i^{-1}.

As the sequences (I) and (II) have the properties (a), (b) and (c), the discussion of the critical points of ϕ_i^{-1} and ψ_i^{-1} in (A) and (B) can be

---

* A, p. 403. If ψ_{i+1} is linear, we have systems of one letter. Of course, such systems are not commonly called systems of imprimitivity.

† A, p. 405, line 26.

‡ Of course, (c) and (b) imply (a). But (a) was not derived from (b) and (c) in paper A, so that the proof of (a) in §8 is not a redundancy.

carried over without modification to the sequences (I) and (II). We are in a position to state, for i sufficiently large, how the branches of ϕ_i^{-1} and ψ_i^{-1} in (I) and (II) are permuted at their critical points.

In what follows, we shall consider in turn the several possibilities for ϕ_i and ψ_i, consulting A at each step. But we must first stop to consider the case in which, for i sufficiently large, the functions in one of the sequences (I) and (II) are linear.

10. **Case** (b). Suppose that, for some i, one or both of the functions ϕ_i, ψ_i are linear. There is no generality lost in assuming, when ϕ_i and ψ_i are not both linear, that ϕ_i is linear rather than ψ_i. If necessary, we can study the pair of functions $f(z)$ and $f[\lambda^{-1}(z)]$, thus interchanging the rôles of (I) and (II). We suppose then that ϕ_i is linear.

As we have a linear function at our disposal in choosing $f_{i+1}(z)$, we may evidently suppose that $\phi_i(z)=z$, so that $f_{i+1}(z)=f_i(z)$, and

$$f_i[\lambda(z)] = \psi_i[f_i(z)]. \tag{4}$$

In the case of $\lambda(z)=mz$, $f_i(z)$ is a meromorphic function with a *rational multiplication theorem*. Such functions are so thoroughly described by an existence theorem of Poincaré,* that they may be regarded as completely determined.

In stating this theorem, we shall assume that $f_i(0)=0$. This can always be brought about by replacing $f_i(z)$ by a linear function of itself. It is easy to prove that, in a relation like (4), $\psi_i'(0)$ must be an integral power of m. Poincaré shows that, given any rational $\psi_i(z)$ with $\psi_i(0)=0$ and $|\psi_i'(0)|>1$, and given an m such that $m^n=\psi_i'(0)$ (n a positive integer), there exists a meromorphic $f_i(z)$ with n zeros at the origin which satisfies (4). Every other solution of (4) is obtained from this particular solution by replacing z by a constant times z.

In a relation (4) with $|m|\leqq 1$, $\psi_i(z)$ must be linear. Relinquishing the condition that $f_i(0)=0$, we may, replacing $f_i(z)$ by a suitable linear function of itself, so arrange matters that $\psi_i(z)$ is of the form $pz+q$. The discussion of $f_i(z)$ then becomes analogous to that of $g(z)$ in §4.

The case of $\lambda(z)=z+m$ leads to an $f_i(z)$ with a *rational addition theorem*. Thus far, no attempt seems to have been made to determine all such functions. Given a function with a rational multiplication theorem, we can obtain from it a function with an addition theorem by replacing z by e^z, or, more generally, by $e^z\pi(z)$, where $\pi(z)$ is a periodic function with a suitable period.

---

* Loc. cit.

But it is far from certain that all functions with rational addition theorems can be found in this way.

Given any $f_i(z)$ such that $f_i[\lambda(z)]$ is rational in $f_i(z)$, any $f(z)$ which is rational in $f_i(z)$ satisfies with $f[\lambda(z)]$ a relation of genus zero. It is the functions $f(z)$ of this type which give Case (b) under genus zero in the statement of results in the introduction.

11. **Preliminaries for Case** (c). We start now our investigation of the cases in which the sequences (I) and (II) contain no linear functions. We take an i_1 so great that for every $i \geqq i_1$, ϕ_i^{-1} has r branches and $h \leqq 4$ critical points, while ψ_i^{-1} has s branches and $k \leqq 4$ critical points.

We shall work under the assumption that $i_1 = 0$. That is, $f_{i_1}(z)$ will become a new $f_0(z)$. At the end of our work we shall replace $f_0(z)$ by an arbitrary rational function of $f_0(z)$, thus allowing for the functions $\phi_1, \cdots, \phi_{i_1-1}$ which will have been suppressed.

We assume also that $r \geqq s$. This amounts to replacing, if necessary, $f_0(z)$ and $f_0[\lambda(z)]$ by $f_0(z)$ and $f[\lambda^{-1}(z)]$. When neither h nor k exceeds 3, we assume further that $h \geqq k$.*

12. **Case of** $h(z) = e^z$. We take the case in which $h = k = 2$, and in which r and s are not both 2. In this case, ϕ_i^{-1} and ψ_i^{-1} have, for every i, the same two critical points, a_i and b_i, at each of which their branches are permuted in a single cycle. The two values assumed by ϕ_i^{-1} and ψ_i^{-1} at their critical points are a_{i+1} and b_{i+1}, the critical points of ϕ_{i+1}^{-1} and ψ_{i+1}^{-1}.

As we can use any linear function of $f_i(z)$ instead of $f_i(z)$ with the same effect, we may replace ϕ_i and ψ_i by any linear function of themselves. We therefore assume, without loss of generality, that $a_i = 0$, $b_i = \infty$ for every i.

It follows immediately that $\phi_i(z) = c_i z^{\pm r}$, where c_i is a constant. Hence

$$f_1(z) = C_i[f_i(z)]^{\pm r^{i-1}}, \tag{5}$$

with C_i a constant.

We infer that $f_1(z)$ is an integral function devoid of zeros. For if $f_1(z)$ had a pole or a zero at z_1, then $f_i(z)$ $(i > 0)$ would also have a pole or a zero at z_1. By taking i large, we could, using (5), make the order of the pole or zero of $f_1(z)$ appear arbitrarily large.

Thus $f_1(z) = e^{g(z)}$, where $g(z)$ is an integral function. Then, since

$$\phi_0(z) = c_0 z^{\pm r}, \qquad \psi_0(z) = d_0 z^{\pm s},$$

* A, p. 403. The arrangement of the assumptions with respect to r and s, h and k, is different in this paper from that in A. In A it was assumed that $h \geqq k$. This is not compatible, when $h = 4$, with the assumption that $r \geqq s$. It is desirable in this paper to have $r \geqq s$ when one of the numbers h or k is 4. We shall still be able to take over from A the discussion of the case of four critical points.

with c_0 and d_0 constants, we have

$$f_0(z) = c_0 e^{\pm rg(z)}, \qquad f_0[\lambda(z)] = d_0 e^{\pm sg(z)}.$$

But as $f_0[\lambda(z)] = c_0 e^{\pm rg[\lambda(z)]}$, we must have

$$g[\lambda(z)] = pg(z) + q, \tag{6}$$

with $p = \pm s/r$ and $q = (\pm 1/r) \log (c_0/d_0)$.

Thus, when $h=k=2$, $f(z)$* is a rational function of $e^{g(z)}$, where $g(z)$ is an integral function satisfying (6).

Any $f(z)$ which is rational in $e^{g(z)}$, where $g(z)$ obeys (6) for a rational p, and any q, will satisfy with $f[\lambda(z)]$ a relation of genus zero. It can be shown that if $f(z)$ is not a constant, and if neither p nor $1/p$ is integral, no ϕ_i or ψ_i will be linear. But not always will h equal 2. For $f(z) = \cos z$, for instance, as will be seen in the following section, $h=3$. This state of affairs is due to the fact that every $f_{i+1}(z)$ is determined to within a linear function when $f_i(z)$ is known, so that, if $f(z) = R[e^{g(z)}]$, with $R(z)$ rational, $R(z)$ influences the formation of the sequences (I) and (II).

13. **Case of $h(z) = \cos z$.** In all cases for which $h=3$, we denote the critical points of ϕ_i^{-1} by a_i, b_i, c_i, and the orders of the substitutions which the branches of ϕ_i^{-1} undergo at a_i, b_i, c_i, by α_i, β_i, γ_i respectively.

We consider now the case in which, for every i, $\alpha_i = \beta_i = 2$. It is proved in *A* that, at a_i and b_i, the branches of ϕ_i^{-1} and ψ_i^{-1} are permuted in pairs, except that there are two places on the Riemann surfaces of ϕ_i^{-1} and ψ_i^{-1} at a_i and b_i where the functions are uniform. At c_i, the branches of the two functions are permuted in single cycles. At the places at a_i and b_i at which they are uniform, ϕ_i^{-1} and ψ_i^{-1} assume the same pair of values, a_{i+1} and b_{i+1}, which are critical points of ϕ_{i+1}^{-1} with $\alpha_{i+1} = \beta_{i+1} = 2$. Also, at c_i, ϕ_i^{-1} and ψ_i^{-1} have a common value, c_{i+1}, which is the third critical point of ϕ_{i+1}^{-1}.

It is easy to see that the branches of the inverse of

$$F_i = \phi_0 \, \phi_1 \cdots \phi_i$$

are permuted in a single cycle at c_0, that is, $F_i(z)$ assumes the value c_0 only for $z = c_i$, and assumes it there r^{i+1} times. It follows that $f_0(z)$ never assumes the value c_0. For $f_0(z)$ could equal c_0 only where $f_i(z)$ would equal c_i, and hence would have to assume the value c_0 an arbitrarily large number of times wherever it assumed it. Similarly no $f_i(z)$ ever equals c_i.

---

* $f(z)$ is the rational function of $f_0(z)$ mentioned in §11.

We use the linear functions at our disposal in selecting the functions $f_i(z)$ to throw each c_i to ∞.* Then every $f_i(z)$ is an integral function.

Suppose that wherever $f_1(z)$ assumes the values a_1 and b_1, it assumes them an even number of times. We use the two constants still at our disposal in choosing $f_1(z)$ to arrange so that $a_1=1$, $b_1=-1$.

We consider now the function arccos z. In the finite part of the plane it has just two critical points, 1 and -1, and its branches are permuted in pairs at those points. Also arccos z is never infinite. Hence the expression arccos $f_1(z)$ represents an infinite number of integral functions. Let $j(z)$ be one of these. Then $f_1(z)=\cos j(z)$.

We have, a fortiori, $f_1(z)=\cos (g(z))^{1/2}$, where $g(z)$ is an integral function.

We lay this result aside for the present, and examine the more difficult case in which one of the values a_1 and b_1, say a_1, is assumed an odd number of times at some point z_1.

There must be a place on the surface of ϕ_1^{-1} at a_1 where ϕ_1^{-1} is uniform and assumes the value $f_2(z_1)$. The value $f_2(z_1)$ of ϕ_1^{-1} at this place is the affix of a critical point of ϕ_2^{-1}. As we are free to interchange the letters a_i and b_i if necessary, we assume that $f_2(z_1)=a_2$, assigning the letter b_2 to the second finite critical point of ϕ_2^{-1}. Similarly we let $a_i=f_i(z_1)$ for every i.

We now arrange matters so that $a_i=1$, $b_i=-1$ for every $i\geqq 1$.

We have

$$f_{i+1}(z) = F_i^{-1}[f_1(z)],$$

with $F_i=\phi_1 \cdots \phi_i$. The values of $f_{i+1}(z)$ in the neighborhood of z_1 come from that branch of F_i^{-1} which is uniform at $z=1$ and assumes there the value 1.

We shall study that branch of F_i^{-1}. Let $n=r^i$ be the degree of $F_i(z)$. As $\cos z$ assumes the values 1 and -1 twice wherever it assumes them, there exist n integral functions $F_i^{-1}(\cos z)$. Let $\alpha(z)$ be that one of these integral functions which equals 1 for $z=0$. As $\alpha(z)$ assumes the values 1 and -1 only through uniform branches of F_i^{-1}, it assumes those values twice wherever it assumes them. Hence there are an infinite number of integral functions arccos $\alpha(z)$. It is proved in A that these are all linear. Because arccos z assumes the value 0 at one of its branch points at $z=1$, there are two of the linear functions arccos $\alpha(z)$, negatives of each other, which vanish for $z=0$. Let tz be one of these. Then

$$\cos z = F_i(\cos tz),$$

so that, because $F_i(z)$ is of degree n, we have $t=\pm 1/n$. As we may replace t by $-t$, we assume that $t=1/n$.

* No difficulty arises when $i=0$; we may also replace $f_0(z)$ by a linear function of itself.

Thus that branch of F_i^{-1} which equals 1 for $z=1$ has, for the neighborhood of $z=1$, the representation

$$F_i^{-1}(z) = \cos\frac{\arccos z}{n}.$$

It is understood here that we use one of those two branches of arccos z which equal 0 for $z=1$.

Consider the integral function

$$\zeta(z) = \cos z^{1/2}.$$

We have, for the neighborhood of $z=1$,

$$F_i^{-1}(z) = \zeta\left[\left(\frac{\arccos z}{n}\right)^2\right]. \tag{7}$$

As the two branches of arccos z which vanish for $z=1$ are negatives of each other, the square of either of these branches is uniform at $z=1$. If z describes a path beginning at 1, and if the two members of (7) are followed along this path, (7) will remain an equation along this path.

We consider now the function

$$[\arccos f_1(z)]^2. \tag{8}$$

We shall prove that, if at $z=z_1$, where $f_1(z)$ assumes the value 1 an odd number of times, we operate on $f_1(z)$ with that branch of $(\arccos z)^2$ which is uniform at $z=1$, the function (8) obtained is an integral function.

Reject from the plane of z those points other than z_1 at which $f_1(z)=\pm 1$. Let C be the region left after this rejection. Let (8) be continued from z_1 to any point z' in C along a path lying in C. While z follows this path, $f_1(z)$ describes a path from 1 to $f(z')$. The function (7) is analytic for all values of z along this path joining 1 to $f_1(z')$.

Now, by (7), we have, since $F_i^{-1}[f_1(z)]=f_{i+1}(z)$,

$$f_{i+1}(z) = 1 - \frac{[\arccos f_1(z)]^2}{2!n^2} + \frac{[\arccos f_1(z)]^4}{4!n^2} + \cdots$$

or

$$n^2[f_{i+1}(z) - 1] = -\frac{[\arccos f_1(z)]^2}{2!} + \frac{[\arccos f_1(z)]^4}{4!n^2} + \cdots. \tag{9}$$

Suppose now that for two distinct paths in C, leading from z_1 to z', the function $[\arccos f_1(z)]^2$ gives two values, V_1 and V_2, at z'. We let i

increase indefinitely, so that $n=r^i$ also approaches ∞. All terms other than the first in the second member of (9) will approach 0, so that, when i is large, the second member will be close to $-V_1/2$ for the first path and close to $-V_2/2$ for the second. But, for every i, the first member of (9) is a uniform function. Hence $[\arccos f_1(z)]^2$ is uniform (and analytic) in C.

When $f_1(z)=\pm 1$, $[\arccos f_1(z)]^2$ cannot have a singularity if it is uniform, for the branches of $\arccos z$ are all finite for $z=\pm 1$. Thus $[\arccos f_1(z)]^2$ is an integral function.

The net result of our discussion is that, whether $f_1(z)$ always assumes the values ± 1 an even number of times, or whether it assumes the value 1 an odd number of times at some point z_1, we have

$$f_1(z) = \cos (g(z))^{1/2}$$

with $g(z)$ integral. We arrange so that $\phi_0(1)=1$, $\phi_0(-1)=-1$, $\phi_0(\infty)=\infty$, and find that

$$f_0(z) = \cos r(g(z))^{1/2}, \qquad f_0[\lambda(z)] = \cos [s(g(z))^{1/2} + u],$$

where $u=0$ if $\psi_0(1)=1$ and $u=\pi$ if $\psi_0(1)=-1$. Hence

$$r(g[\lambda(z)])^{1/2} = \pm [s(g(z))^{1/2} + u] + 2t\pi$$

with t an integer. We have thus

$$(g[\lambda(z)])^{1/2} = p(g(z))^{1/2} + q \tag{10}$$

where $p=\pm s/r$ and where $q=n\pi/r$ with n integral.

An integral $g(z)$ which satisfies (10) with p and q as just described (r, s, and n are arbitrary) will give an $f_0(z)=\cos (g(z))^{1/2}$ which satisfies with $f_0[\lambda(z)]$ a relation of genus zero. Any $f(z)$ which is rational in $f_0(z)$ satisfies with $f[\lambda(z)]$ a relation of genus zero. If neither p nor $1/p$ is an integer, no ϕ_i or ψ_i will be linear.

If q is chosen different from zero, then $(g(z))^{1/2}$ will be integral. For suppose that $(g(z))^{1/2}$ is not uniform. Then it vanishes at its branch points. As $(g[\lambda(z)])^{1/2}$ has the same branch points as $g(z)$, and vanishes at those points, we must have $q=0$.

When $g(z)$ is integral, it is described as in §4. If we take $q=0$, any integral $g(z)$ for which $g[\lambda(z)]=p^2g(z)$ can be used, whether its square root is integral or not.

14. **Case of $h(z)=\varphi(z)$.** We undertake now the investigation of the following two cases:

(α) Either h or k equals 4. (It is assumed that $r\geqq s$.)

(β) $h=k=2$.

When $h=4$ or $k=4$, the branches of ϕ_i^{-1} and ψ_i^{-1} are permuted in pairs at four certain points, except that there are four places on the surface of each function, at those points, where the functions are uniform, the values of ϕ_i^{-1} and ψ_i^{-1} at these places being the critical points of ϕ_{i+1}^{-1} and ψ_{i+1}^{-1}.

When $h=k=2$, we have $r=s=2$, and the critical points of ϕ_i^{-1} are two points a_i and b_i, while those of ψ_i^{-1} are a_i and a third point $c_i \neq b_i$. The values of ϕ_i^{-1} at c_i are a_{i+1} and c_{i+1} and the values of ψ_i^{-1} at b_i are a_{i+1} and b_{i+1}. There is a point d_i at which one value of ϕ_i is b_{i+1} and at which one value of ψ_i^{-1} is c_{i+1}. The second values of ϕ_i^{-1} and ψ_i^{-1} at d_i are both d_{i+1}.

When h or k equals 4, we represent the critical points of ϕ_i^{-1} and ψ_i^{-1} by a_i, b_i, c_i, and d_i. This will permit us to treat cases (α) and (β) together.

Suppose that $f_1(z)$ assumes the values a_1 etc. an even number of times wherever it assumes them. Replacing $f_1(z)$ by a linear function of itself if necessary, we suppose that $d_1=\infty$, and that $a_1+b_1+c_1=0$. We now construct $\wp(z)$ with $\wp(\omega_1)=a_1$, $\wp(\omega_2)=b_1$, $\wp(\omega_3)=c_1$. We find that

$$f_1(z) = \wp[j(z)]$$

with $j(z)$ an integral function.* Certainly, then,

$$f_1(z) = \wp([g(z)]^{1/2}),$$

where $g(z)$ is an integral function.

We now take the case in which $f_1(z)$ assumes one of the values a_1 etc. an odd number of times at some point z_1. If $h=2$, we must have $f_1(z_1)=d_1$, for the relation

$$f_1[\lambda(z)] = \phi_1(f_2[\lambda(z)]) = \psi_1[f_2(z)]$$

shows that $f(z)$ assumes the values a_1, b_1, c_1 an even number of times whereever it assumes them. In general, we have $f_i(z_1)=d_i$.

For $h=4$ or $k=4$, no special assignment of the letters a_i etc. to the critical points has been made. We do this now.

We let $d_1=f_1(z_1)$, and assign the letters a_1, b_1, c_1 in any way at all to the other critical points. Then there must be a place at d_1 at which ϕ_1^{-1} is uniform and assumes the value $f_2(z_1)$. This value of ϕ_1^{-1} is one of the critical points of ϕ_2^{-1} or ψ_2^{-1}. Denoting this critical point by d_2, we assign the letters a_2, b_2, c_2 arbitrarily. We repeat this procedure indefinitely, putting $d_i=f_i(z_1)$ for every i.

Resuming the simultaneous treatment of Cases (α) and (β), we suppose that, for every $i \geqq 1$, $d_1=\infty$ and $a_i+b_i+c_i=0$. This means, because $\phi_i(z)$

---

* *A*, p. 425.

assumes the value d_i once at d_{i+1}, that each $\phi_i(z)$ has a simple pole at ∞. We may (and shall) suppose further that the derivative of $\phi_i(z)$ equals unity for $z=\infty$. This uses up, for $i>1$, the third constant in the linear transformation at our disposal in choosing $f_i(z)$.

We construct, for each $i\geqq 1$, the function $\varphi(z)$ with such periods that its derivative vanishes when $\varphi(z)$ is a_i, b_i or c_i. We represent this function by $\varphi_i(z)$.

We shall prove that $\varphi_i(z)=\phi_i[\varphi_{i+1}(z)]$. If we operate on $\varphi_i(z)$ with ϕ_i^{-1}, we get r elliptic functions, which assume the values ∞, a_{i+1}, b_{i+1}, c_{i+1} twice wherever they assume them. Because $\phi_i^{-1}(z)$ has a branch with a simple pole and with derivative unity at ∞, one of these elliptic functions has a double pole for $z=0$ with $1/z^2$ for the principal part in its development. Call it $\beta(z)$. Let $\varphi_{i+1}^{-1}(z)$ represent the inverse of $\varphi_{i+1}(z)$. Then there are an infinitude of linear integral functions $\varphi_{i+1}^{-1}[\beta(z)]$.*

But φ_{i+1}^{-1} has one branch point at ∞ where its development begins with $z^{-1/2}$. One of the two corresponding developments of $\varphi_{i+1}^{-1}[\beta(z)]$ for $z=0$ will start with z, and will therefore actually be z. This proves that $\varphi_i(z)=\phi_i[\varphi_{i+1}(z)]$.

Our next step will be to prove that, as i increases, the periods of $\varphi_i(z)$ increase without limit; that is, that given any number G, we can find an integer j such that, for every $i>j$, the periods of $\varphi_i(z)$ are all at least equal in modulus to G.

Let a G be taken. For i sufficiently large, $\varphi_i(z)$ cannot have two periods with a complex ratio whose moduli are both less than G. For in a parallogram based on such periods, $\varphi_i(z)$ would assume all values at least twice, and $\varphi_1(z)$, which is a rational function of degree r^{i-1} in $\varphi_i(z)$, would assume every value at least $2r^{i-1}$ times. If i is large, this produces an absurdity.

Every period of $\varphi_{i+1}(z)$ is also a period of $\varphi_i(z)$. Hence, if there is a $\varphi_i(z)$ with i arbitrarily large, and with a period smaller than G in modulus, there must exist an ω with $|\omega|<G$ which is a period of $\varphi_i(z)$ for every i.

Suppose that this is so. Of the numbers ω which are periods of every $\varphi_i(z)$, we select one, $\bar{\omega}$, of smallest modulus.† It is easy to show that

$$\psi_i[\varphi_{i+1}(z)] = \varphi_i[\epsilon_i z + u_i]$$

where ϵ_i and u_i are constants, $2u_i$ being a period of $\varphi_i(z)$. Because a primitive period parallelogram of $\varphi_{i+1}(z)$ has r times the area of a parallelogram for

---

* *A*, p. 425.

† There will be two such $\bar{\omega}$'s, negatives of each other.

$\varphi_i(z)$ and s times the area of one for $\varphi_i(\epsilon_i z+u_i)$, we must have $|\epsilon_i|^2=s/r$. Hence, as $r\geqq s$, we have $|\epsilon_i|\leqq 1$.

For i large, $\bar{\omega}$ is a smallest period of $\varphi_i(z)$. Then $\bar{\omega}/\epsilon_i$ is a smallest period of $\varphi_i(\epsilon_i z+u_i)$. As $\bar{\omega}$ is a period of $\varphi_i(\epsilon_i z+u_i)$ (it is one of $\varphi_{i+1}(z)$), we must have $|\epsilon_i|\geqq 1$. Hence $|\epsilon_i|=1$.

Thus, unless $\epsilon_i=\pm 1$, $\varphi_i(\epsilon_i z+u_i)$ will have two periods of complex ratio, $\bar{\omega}$ and $\epsilon_i\bar{\omega}$, of moduli less than G. But if $\epsilon_i=\pm 1$, $\varphi_i(\epsilon_i z+u_i)$ will be a linear function of $\varphi_i(z)$, because u_i is a half-period, so that ϕ_i will be a linear function of ψ_i. This contradicts the fact that ϕ_i and ψ_i are not linear and have no common forefactor. The proof that the periods of $\varphi_i(z)$ become infinite with i is thus completed.

Putting $F_i=\phi_1\cdots\phi_i$, we have

$$f_{i+1}(z) = F_i^{-1}[f_1(z)]. \tag{11}$$

For the neighborhood of z_1, we must use, in (11), that branch of F_i^{-1} which equals ∞ at ∞. We proceed to obtain an expression for that branch.

As $\varphi_1(z)=F_i[\varphi_i(z)]$, we have

$$F_i^{-1}(z) = \varphi_i[\varphi_1^{-1}(z)].$$

To get that branch of F_i^{-1} which is infinite at ∞, we must use, for $z=0$, one of those two branches of $\varphi_1^{-1}(z)$ which vanish at ∞.

Let $\zeta_i(z)=\varphi_i(z)^{1/2}$. We have

$$F_i^{-1}(z) = \zeta_i([\varphi_1^{-1}(z)]^2). \tag{12}$$

We use here that branch of $[\varphi_i^{-1}(z)]^2$ which is uniform, and equal to zero, at ∞. Then, by (11) and (12), we have

$$f_{i+1}(z) = \zeta_i([\varphi_1^{-1}\{f_1(z)\}]^2). \tag{13}$$

For the neighborhood of $z=0$, we have

$$\zeta_i(z) = \frac{1}{z} + l_1 z + l_2 z^2 + \cdots. \tag{14}$$

Since the periods of $\varphi_i(z)$ become infinite with i, $\varphi_i(z)$ degenerates towards $1/z^2$ as i increases. Hence the radius of convergence of the development in (14) increases indefinitely with i, while each coefficient l_n approaches zero.

We shall now prove that $(\varphi_1^{-1}[f_1(z)])^2$ is an integral function. We take any finite region about z_1, reject from it all points other than z_1 for which $f_1(z)$ equals ∞, a_1, b_1, or c_1, and let z describe two different paths from z_1 to any point z' in this region. Then, because $\varphi_1^{-1}(z)$ is never infinite,

$(\varphi_1^{-1}[f_1(z)])^2$ will describe two bounded paths, which start from 0 and which terminate at two points V_1 and V_2.

From (13) and (14), we find

$$f_{i+1}(z) = \frac{1}{(\varphi_1^{-1}[f_1(z)])^2} + l_1(\varphi_1^{-1}[f_1(z)])^2 + \cdots . \tag{15}$$

For i large enough, the paths leading from 0 to V_1 and V_2 will both lie within the circle of convergence of (14). As the coefficients l_n approach 0 with $1/i$, and as the first member of (15) is uniform, V_1 and V_2 cannot be distinct. This means that $f_1(z)$ is of the form $\varphi_1[(g(z))^{1/2}]$ where $g(z)$ is an integral function.

At this point we bring together the case in which $f_1(z)$ always assumes the values a_1 etc. an even number of times with that in which it assumes one of the values somewhere an odd number of times. We put d_0 equal to $\phi_0(d_1)$,* and send d_0 to ∞. Also we make $a_0+b_0+c_0$ equal 0. Finally, we make $\phi_0'(\infty)$ unity. Taking a $\varphi(z)$ whose multiple values are a_0 etc., we see that

$$f_0(z) = \varphi[(g(z))^{1/2}], \qquad f_0[\lambda(z)] = \varphi[\epsilon(g(z))^{1/2} + u]$$

where $2u$ is a period of $\varphi(z)$, and ϵ a constant.

Then

$$(g[\lambda(z)])^{1/2} = \pm \epsilon(g(z))^{1/2} + v,$$

where $2v$ is a period of $\varphi(z)$.

The constant ϵ must be such that $\varphi(z)$ and $\varphi(\epsilon z)$ have a pair of periods in common. It will always be the quotient of two numbers ϵ_1 and ϵ_2 such that $\varphi(\epsilon_1 z)$ and $\varphi(\epsilon_2 z)$ are rational functions of $\varphi(z)$. Any such ϵ, and any half-period v can be used for the determination of a $g(z)$. If $v \neq 0$, $g(z)$ must be integral. If $v=0$, we can use any $g(z)$ for which $g[\lambda(z)] = \epsilon^2 g(z)$, whether $(g(z))^{1/2}$ is integral or not.

Any $f(z)$ which is rational in $\varphi(g(z))^{1/2}$ satisfies with $f[\lambda(z)]$ an equation of genus zero.

15. **The lemniscatic functions.** We take the case of $h=3$, $\alpha_i=2$, $\beta_i=\gamma_i=4$.

Suppose first that wherever $f_1(z)$ assumes the value a_1 it assumes it an even number of times, and that where it assumes the values b_1 and c_1 it assumes them a number of times which is divisible by 4.

---

* When $h=k=2$, this relation already holds.

We construct the function $\wp(z|\omega, i\omega)$, where ω is any number distinct from zero. For this function, $e_1 = -e_3$. We arrange so that $a_1 = e_1^2$, $b_1 = 0$, $c_1 = \infty$, and prove that

$$f_1(z) = \wp^2[j(z)]$$

where $j(z)$ is an integral function. Certainly

$$f_1(z) = \wp^2[(g(z))^{1/4}],$$

where $g(z)$ is integral.

Suppose now that, at z_1, $f_1(z)$ assumes one of the values b_1, c_1 a number of times which is not divisible by 4. Let the designation c_1 be assigned to $f_1(z_1)$.

The value $f_2(z_1)$ must be assumed by ϕ_1^{-1} at c_1 through a uniform branch. This is evident if the value c_1 is assumed by $f_1(z)$ at z_1 an odd number of times. If c_1 is assumed at z_1 an even number of times (not divisible by 4), it might be asked whether ϕ_1^{-1} does not assume the value $f_2(z_1)$ at a simple branch point at c_1. Such a simple branch point may exist if r is even.* But then $f_2(z_1)$ would equal a_2.† Also ϕ_2^{-1} would have no uniform branches at a_2, and the value a_2 would be assumed an odd number of times by $f_2(z)$ at z_1. Thus $f_3(z)$ could not be uniform at z_1.

The value $f_2(z_1)$, assumed by ϕ_1^{-1} at a uniform branch at c_1, is a critical point of ϕ_2^{-1} at which the branches of ϕ_2^{-1} undergo a substitution of order 4. We put $f_2(z_1) = c_2$, whereupon the use of the letter b_2 is determined. In general, we put $f_i(z_1) = c_i$.

Taking the function $\wp(z|\omega, i\omega)$, and letting

$$a_i = e_1^2, \quad b_i = 0, \quad c_i = \infty$$

for every $i \geqq 1$, it can be proved, using the method of the preceding sections, that

$$f_1(z) = \wp^2[(g(z))^{1/4}]$$

where $g(z)$ is integral. The work is easier than that for the case of four critical points, because we meet in the present case, as in §13, multiplication theorems rather than period transformations. No discussion similar to that, in the preceding section, of the size of the periods of $\wp_i(z)$, is necessary.

---

* *A*, p. 428.

† *A*, p. 428, line 27.

In both cases treated above,—in the case in which the value c_1 is always assumed a number of times divisible by 4, and in that in which it is not,—we find, putting

$$a_0 = e_1^2, \quad b_0 = 0, \quad c_0 = \phi_0(c_1) = \infty,$$

$$f_0(z) = \wp^2[\epsilon_1(g(z))^{1/4}], \quad f_0[\lambda(z)] = \wp^2[\epsilon_2(g(z))^{1/4} + u].$$

Here ϵ_1 and ϵ_2 are constants such that $\wp^2(\epsilon_1 z)$ and $\wp^2(\epsilon_2 z)$ are rational in $\wp^2(z)$, and u is either 0 or $(1+i)\omega$. As to $g(z)$, we must have

$$\epsilon_1(g[\lambda(z)])^{1/4} = i^p\epsilon_2(g(z))^{1/4} + v$$

where p is either 0, 1, 2 or 3, and where v is a constant easy to specify. If $v \neq 0$, $(g(z))^{1/4}$ is integral.

We consider finally the case in which the value a_1 is assumed by $f_1(z)$ at some point z_1 an odd number of times. We prove as above that

$$f_0(z) = \wp^2[\epsilon_1(g(z))^{1/2} + \omega], \quad f[\lambda(z)] = \wp^2[\epsilon_2(g(z))^{1/2} + u]$$

where u is either ω or $i\omega$.

16. **The equianharmonic functions.** In the case of $h=3$, $\alpha_i=\beta_i=\gamma_i=3$, we construct $\wp(z)$ with $e^{2\pi i/3}$ for the ratio of its periods. We prove, assigning appropriate values to a_0, b_0, c_0, that

$$f_0(z) = \wp'[\epsilon_1(g(z))^{1/3}], \quad f_0[\lambda(z)] = \wp'[\epsilon_2(g(z))^{1/3} + u].$$

Here ϵ_1 and ϵ_2 are such that $\wp'(\epsilon_1 z)$ and $\wp'(\epsilon_2 z)$ are rational in $\wp'(z)$, and $g(z)$ is integral. The constant u is either 0 or a point at which $\wp''(z)=0$. We omit the discussion of $g(z)$.

In the case of $\alpha=2$, $\beta=3$, $\gamma=3$ or 6, we take the function $\wp(z)$ just constructed and show that

$$f_0(z) = \wp^3[\epsilon_1(g(z))^{1/n} + u], \quad f_0[\lambda(z)] = \wp^3[\epsilon_2(g(z))^{1/n} + u],$$

where $n=6$, 3 or 2. If $n=6$, u is 0. If $n=3$, u is such that $\wp(u)=0$, while if $n=2$, $\wp^3(u)=e_1^3$.

In this section, and in the preceding one, any $f(z)$ which is rational in $f_0(z)$ satisfies with $f[\lambda(z)]$ a relation of genus zero.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

REAL FUNCTIONS WITH ALGEBRAIC ADDITION THEOREMS*

BY

J. F. RITT

It is a well known theorem of Weierstrass that if an analytic function $\phi(x)$ satisfies, for all values of x and y, an algebraic relation

$$G[\phi(x),\ \phi(y),\ \phi(x+y)] = 0, \tag{1}$$

the function $\phi(x)$ is either an algebraic function, or an algebraic function of $e^{\mu x}$ (μ constant), or an algebraic function of $\wp(x)$. Weierstrass never published a proof of this theorem. Proofs were given by Phragmen and by Koebe.†

If one restricts oneself to the real domain, it is natural to require less than analyticity of $\phi(x)$, for instance, mere continuity. In that case, $\phi(x)$ no longer has to be a function of one of the three types just mentioned. For example, if $\phi(x)$ is one algebraic function of $\wp(x)$ in an interval (a, b), and another algebraic function of the same $\wp(x)$ in (b, c), $\phi(x)$ will have an algebraic addition theorem.

Even if the polynomial G in (1) is irreducible, $\phi(x)$ need not be analytic. For instance, the addition theorem of $\sin x$ is

$$(w^2 - u^2 - v^2 + 2u^2v^2)^2 - 4u^2v^2(1 - u^2)(1 - v^2) = 0 \tag{2}$$

where

$$u = \sin x, \quad v = \sin y, \quad w = \sin (x + y).$$

The first member of (2) is irreducible.‡ Now, because (2) involves u, v and w only in even powers, (2) also furnishes an addition theorem for the continuous but non-analytic function $|\sin x|$.

In the present paper, we consider functions $\phi(x)$, continuous in an interval $(0, a>0)$, and possessing an addition theorem (1) with G irreducible. We obtain, relative to $\phi(x)$, the three theorems which follow.

---

* Presented to the Society, February 27, 1926; received by the editors in April, 1926.

† Phragmen, Acta Mathematica, vol. 7 (1885), p. 33; Koebe, *Mathematische Abhandlungen H. A. Schwarz zu seinem fünfzigjährigen Doctorjubiläum gewidmet*, p. 192.

‡ This follows from the fact that to given values of $\sin x$ and $\sin y$, there correspond four values of $\sin(x+y)$.

THEOREM I. *Let $\phi(x)$, real and continuous in $(0, a)$ $(a>0)$, satisfy for $0 \leqq x \leqq a$, $0 \leqq y \leqq a-x$, a relation*

$$G[\phi(x), \phi(y), \phi(x+y)] = 0,$$

in which G is an irreducible polynomial in its three arguments. Then $\phi(x)$ is piecewise analytic in $(0, a)$, that is, $(0, a)$ can be divided into a finite number of intervals in the interior of each of which $\phi(x)$ coincides with an analytic function.

Let the intervals in which $\phi(x)$ is analytic be

$$(x_0, x_1), \quad (x_1, x_2), \cdots, (x_{n-1}, x_n) \qquad (x_0 = 0, x_n = a).$$

Let $\phi_i(x)$ be used to represent $\phi(x)$ in (x_{i-1}, x_i). With respect to $\phi_2, \cdots, \phi_n$, we prove, with the help of a method due to Koebe, the following two theorems:

THEOREM II. *For every $i \leqq n$, there exists a constant α_i such that*

$$\phi_i(x) = \phi_1(x + \alpha_i). \tag{3}$$

THEOREM III. *If there is a point in (x_{k-1}, x_k) which is the sum of a point in (x_{i-1}, x_i) and a point in (x_{j-1}, x_j), then*

$$\phi_i(x) = \phi_1(\epsilon x + d_1), \quad \phi_j(x) = \phi_1(\epsilon x + d_2), \quad \phi_k(x) = \phi(\epsilon x + d_1 + d_2)$$

where d_1 and d_2 are constants, and where $\epsilon = \pm 1$ unless $\phi_1(x)$ is an algebraic function of $\wp(x)$ with $g_3 = 0$ or $g_2 = 0$. If $\phi_1(x)$ is algebraic in $\wp(x)$ with $g_3 = 0$, ϵ may be $\pm(-1)^{1/2}$, whereas, if $g_2 = 0$, ϵ may be a sixth root of unity.*

Given a $\phi_1(x)$ with an addition theorem, the determination of all functions $\phi(x)$ continuous in $(0, a)$ which have the same addition theorem as $\phi_1(x)$ depends on the determination of all possibilities for the constants ϵ and d in Theorem III.

PROOF OF THEOREM I

1. Let $\phi(x)$ satisfy the hypothesis of Theorem I. If $\phi(x)$ is a constant, Theorem I is certainly true for it. Suppose that $\phi(x)$ is not a constant. We shall show first that there exists an interval in which $\phi(x)$ assumes no value twice, and, after this, that there is an interval in which $\phi(x)$ is analytic. It will then be easy to complete the proof of our theorem.

2. We shall study the circumstances under which $\phi(x)$ assumes equal values at two points, x and $x+y$. If, in G, we put $\phi(x+y) = \phi(x)$, we obtain

* The quantities ϵ, d_1, d_2 depend on i, j, k.

a polynomial $H[\phi(x), \phi(y)]$. Here H is not identically zero. If it were, G would be divisible by $\phi(x+y)-\phi(x)$. Then either G would be reducible, or G would be a constant times $\phi(x+y)-\phi(x)$. In the latter case, $\phi(x)$ would be a constant.

Because $H=0$ when $\phi(y)=\phi(0)$, H is divisible by $\phi(y)-\phi(0)$. Let

$$H = [\phi(y) - \phi(0)]^n K[\phi(x), \phi(y)], \tag{4}$$

where K is a polynomial not divisible by $\phi(y)-\phi(0)$. Then the equation $K[\phi(x), \phi(0)]=0$ has only a finite number of solutions for $\phi(x)$.

We can thus find a c in $(0, a)$ such that $K[\phi(c), \phi(0)]\neq 0$. Then if x is close to c, and y close to 0, $K[\phi(x), \phi(y)]$ is not 0.

Suppose that there does not exist an infinite sequence of y's, decreasing towards 0, for each of which $\phi(y)=\phi(0)$. We see from (4), in that case, that for x close to c the equation $H=0$ will not be satisfied by a y close to, and distinct from 0. We have thus an interval about c in which $\phi(x)$ assumes no value more than once.

We shall show that, because $\phi(x)$ is not a constant, there cannot be a sequence of y's decreasing towards 0 for each of which $\phi(y)=\phi(0)$. Suppose that there does exist such a sequence of positive values of y,

$$\alpha_1, \alpha_2, \cdots, \alpha_n, \cdots \tag{5}$$

with $\phi(\alpha_n)=\phi(0)$ $(n=1, 2, \cdots)$, and $\lim \alpha_n=0$. As G is irreducible, and not equal to a constant times $\phi(x)-\phi(0)$, it cannot vanish identically when $\phi(x)$ is replaced by $\phi(\alpha_n)=\phi(0)$. We consider the polynomial

$$G[\phi(\alpha_n), \phi(y), \phi(x+y)]. \tag{6}$$

Because $\phi(x)$ assumes an infinite number of values in $(0, a)$, there is an $h=\phi(\beta)\neq\phi(a)$ $(0\leqq\beta<a)$ such that (6) is not divisible by $\phi(y)-h$. Then the equation

$$G[\phi(\alpha_n), h, \phi(x+y)] = 0 \tag{7}$$

is satisfied by only a finite number of values of $\phi(x+y)$. We shall suppose that β is the greatest value of x less than a for which $\phi(x)=h$. Then (7), when solved for $\phi(x+y)$, will give all values which $\phi(x)$ can possibly assume at the points $\alpha_n+\beta$. As there are only a limited number of solutions of (7), there is an infinite set of points among the points $\alpha_n+\beta$ at which $\phi(x)$ assumes only a single value. Because $\phi(x)$ is continuous, this single value must be $\phi(\beta)$. This contradiction of the assumption that β is a maximum proves that no sequence (5) exists. This, as we have seen, implies the

existence of a point c in the neighborhood of which $\phi(x)$ assumes no value more than once.

3. Thus, in some interval about c, $\phi(x)$ is monotone. A well known theorem of Lebesgue tells us that $\phi(x)$ has a finite derivative almost everywhere in this interval.

This information secured, we proceed to show the existence of an interval in which $\phi(x)$ is analytic.

4. If the equation $G=0$ is of degree n in $\phi(x+y)$, it determines $\phi(x+y)$ as an algebraic function of n branches in $\phi(x)$ and $\phi(y)$. There must be a γ in the interval about c found above, and a δ close to 0, such that the n branches of $\phi(x+y)$ are analytic, and distinct from one another, at $[\phi(\gamma), \phi(\delta)]$. This is because $\phi(x)$ is not constant in the neighborhoods of c and 0. Considered as functions of x and y, the n branches will be continuous functions, distinct from one another, for x near γ and y near δ. Consequently $\phi(x+y)$ is identical with one of these functions for x near γ and y near δ.

We may evidently suppose γ to have been so chosen that $\phi(x)$ has a finite derivative at $x=\gamma$. This understood, we write, for x and y close to γ and δ respectively,

$$\phi(x+y) = F[\phi(x), \phi(y)], \tag{8}$$

where F is analytic (and algebraic) in its arguments. Let y be given any value close to δ, and let x vary past γ. Then the second member of (8) becomes a function of x with a finite derivative for $x=\gamma$. We infer from (8) that $\phi(x)$ has a finite derivative for every x close to $\gamma+\delta$, or, what is the same, that $\phi(\gamma+y)$ has a finite derivative with respect to y for y close to δ.

For $x=\gamma$, and y near δ, we have, from (8),

$$\frac{d}{dx}\phi(x+y) = \frac{d}{dy}\phi(\gamma+y) = U[\phi(\gamma), \phi(y)]\phi'(\gamma), \tag{9}$$

where U is the partial derivative of F with respect to $\phi(x)$. Also

$$\phi(\gamma+y) = F[\phi(\gamma), \phi(y)]. \tag{10}$$

We suppose now that δ is taken so small that $\gamma+\delta$ lies in the interval about c in which $\phi(x)$ is monotone. Then, as y varies past δ, $\phi(\gamma+y)$ will increase or decrease monotonically. Hence, there is a value η of y, close to δ, at which the derivative of $\phi(\gamma+y)$ with respect to y is not 0. Referring to (10), we see that the derivative of $F[\phi(\gamma), \phi(y)]$ with respect to $\phi(y)$ cannot be 0 at $\phi(\eta)$. We can thus solve for $\phi(y)$ in (10), expressing it as an

algebraic function of $\phi(\gamma+y)$, analytic for $\phi(\gamma+y)$ close to $\phi(\gamma+\eta)$. If we substitute this expression for $\phi(y)$ into (9), we have

$$\frac{d}{dy}\phi(\gamma+y) = V[\phi(\gamma+y)], \tag{11}$$

where V is analytic for $\phi(\gamma+y)$ close to $\phi(\gamma+\eta)$.

The existence theorems for differential equations inform us that, because $\phi(\gamma+y)$ satisfies (11), it is analytic for $y=\eta$. Thus $\phi(x)$ is analytic for x close to $\gamma+\eta$.

5. We shall now complete the proof of Theorem I. Consider an interval (α, β), interior to $(0, a)$, in which $\phi(x)$ is monotone and analytic. The existence of such an interval was just proved. Let γ and δ be any numbers in $(0, a)$ such that $\delta-\gamma=\beta-\alpha$. We shall show that (γ, δ) can be divided into a finite number of intervals in each of which $\phi(x)$ coincides with some analytic function. This will, of course, settle the proof of Theorem I.

Suppose first that $\gamma \leqq \alpha$. For x in (γ, δ) we have

$$G[\phi(x), \phi(\alpha-\gamma), \phi(u)] = 0, \tag{12}$$

where $u=x+\alpha-\gamma$. As x ranges over (γ, δ), u ranges over (α, β).

Because G is irreducible, the first member of (12) is not identically zero. Hence, except for a finite number of values of $\phi(u)$, for which (12) may hold for every $\phi(x)$, (12) determines $\phi(x)$ as one of several algebraic functions of $\phi(u)$. Let $\phi(\alpha)=s$, $\phi(\beta)=t$. We shall study $\phi(x)$ as a function of $\phi(u)$ as $\phi(u)$ varies from s to t, that is, since $\phi(u)$ is monotone in (α, β), as u increases from α to β.

To fix our ideas, suppose that $s<t$. Each of the algebraic functions of $\phi(u)$ defined by (12) has only a finite number of singular points. These singular points divide (s, t) into a finite number of intervals in which each branch of each algebraic function is analytic. Consider one of these intervals. In it, two branches of the same or of different functions can be equal at only a finite number of points. Hence we can divide (s, t) into a finite number of intervals in each of which (12) defines a certain number of analytic functions of $\phi(u)$, no two functions being equal at any point interior to an interval.

Let (s', t') be one of these intervals, and let (α', β') be the corresponding interval for u. As u varies over (α', β'), (12) gives a certain number of functions of u, analytic, and nowhere equal to one another, in the interior of (α', β'). One of these functions must coincide with $\phi(x)$ for x in $(\gamma+\alpha'-\alpha, \gamma+\beta'-\alpha)$.

Thus $\phi(x)$ is analytic for x interior to $(\gamma+\alpha'-\alpha,\ \gamma+\beta'-\alpha)$, which is any one of a certain finite number of intervals into which our procedure has divided $(\gamma,\ \delta)$.

A similar proof holds when $\gamma>\alpha$. Theorem I is established.

Proofs of Theorems II and III

6. Let the points $x_0=0, x_1>x_0, \cdots, x_n=a$ divide $(0, a)$ into n intervals, in the interior of each of which $\phi(x)$ is analytic. Let the analytic function with which $\phi(x)$ coincides in (x_{i-1}, x_i) be represented by $\phi_i(x)$. Because $\phi_1(x)$ has the algebraic addition theorem

$$G[\phi_1(x),\ \phi_1(y),\ \phi_1(x+y)] = 0, \tag{13}$$

$\phi_1(x)$ is an algebraic function, or an algebraic function of $e^{\mu x}$, or an algebraic function of $\wp(x)$. As each $\phi_i(x)$ is algebraically related to $\phi_1(x)$, each $\phi_i(x)$ is also of one of the above types, and, indeed, it is of the same type as $\phi_1(x)$.

We are going to show that Theorems II and III hold for the functions $\phi_i(x)$. The method we shall use is one which appears in Koebe's paper, where it serves a different purpose.

7. Let ξ, η, and $\xi+\eta$ be three points situated respectively in (x_{i-1}, x_i), (x_{j-1}, x_j), (x_{k-1}, x_k), at each of which $\phi_1(x)$ has at least one branch which is analytic. Whether or not the branches are real is immaterial. It is proved by Koebe* that no matter which branches are chosen at ξ, η and $\xi+\eta$, equation (13) will be satisfied for x close to ξ and y close to η.

As G is irreducible, and as the equations (13) and

$$G[\phi_i(x),\ \phi_j(y),\ \phi_k(x+y)] = 0$$

are satisfied for x near ξ and y near η, it is possible to vary $\phi_1(x)$ in (13) continuously from $\phi_1(\xi)$ to $\phi_i(\xi)$ and $\phi_1(y)$ from $\phi_1(\eta)$ to $\phi_j(\eta)$, in such a way that $\phi_1(x+y)$ varies from $\phi_1(\xi+\eta)$ to $\phi_k(\xi+\eta)$. These variations will occur while x varies from ξ to a point ξ_1, y from η to a point η_1, and hence, by (13), as $x+y$ varies from $\xi+\eta$ to $\xi_1+\eta_1$. The points ξ_1, η_1 need not be real.

We have

$$\phi_1(\xi_1) = \phi_i(\xi),\quad \phi_1(\eta_1) = \phi_j(\eta),\quad \phi_1(\xi_1+\eta_1) = \phi_k(\xi+\eta).$$

Changing ξ and η slightly if necessary, we may suppose that $\phi_1(x)$ is analytic, with a non-vanishing derivative at ξ_1, η_1, $\xi_1+\eta_1$, and that the equation

$$G[\phi_1(\xi_1),\ \phi_1(\eta_1),\ \phi_1(x+y)] = 0 \tag{14}$$

for $\phi_1(x+y)$ has no equal roots.

* Loc. cit., §3.

We associate with every x close to ξ that x_1 close to ξ_1 for which $\phi_1(x_1)=\phi_i(x)$. With every y close to η we associate that y_1 close to η_1 for which $\phi_1(y_1)=\phi_j(y)$. Thus, because (14) has no equal roots, every $x+y$ close to $\xi+\eta$ is associated with an x_1+y_1 close to $\xi_1+\eta_1$ such that $\phi_1(x_1+y_1)=\phi_k(x+y)$. Let

$$x_1 = u(x), \quad y_1 = v(y), \quad x_1 + y_1 = w(x + y).$$

Here u, v and w are analytic. Then

$$w(x + y) = u(x) + v(y). \tag{15}$$

It is easy to prove, differentiating (15), that

$$u(x) = cx + d_1, \quad v(y) = cy + d_2,$$

where the coefficients are constants. Hence

$$\phi_i(x) = \phi_1(cx + d_1), \quad \phi_j(x) = \phi_1(cx + d_2), \quad \phi_k(x) = \phi_1(cx + d_1 + d_2). \tag{16}$$

Consider any $k \leqq n$. Let a ξ be chosen in (x_0, x_1), and then let an η be taken in such a way that $\xi+\eta$ lies in (x_{k-1}, x_k). Then, in (16), $i=1$, and the first and third relations give

$$\phi_k(x) = \phi_1\left(x + \frac{d_2}{c}\right). \tag{17}$$

This proves Theorem II.

8. We are going to prove, in this section and in §9, that c in (16) is a root of unity, which is ± 1 unless the functions $\phi_i(x)$ are algebraic functions of $\wp(x)$ with $g_3=0$ or $g_2=0$. In the exceptional cases, c may be $\pm i$ if $g_3=0$, whereas, if $g_2=0$, c may be a sixth root of unity.* Theorem III will be established as soon as these facts are proved.

From (3) and (16) we find, replacing x by $x-\alpha_j$, that

$$\phi_1(x) = \phi_1(cx - c\alpha_j + d_2). \tag{18}$$

If, supposing that $c \neq 1$, we put $\beta=(-c\alpha_j+d_2)/(1-c)$, (18) becomes

$$\phi_1(x) = \phi_1[c(x - \beta) + \beta]. \tag{19}$$

This shows that c is a root of unity, for, if it were not, $\phi_1(x)$ would assume values infinitely often in the neighborhood of β. For roots of unity of any order, there are *algebraic* functions which satisfy relations like (19). But if $\phi_1(x)$ is *periodic*, $c=\pm 1$ except in the two cases mentioned above.†

* Similar details occur in Koebe's paper. What is new here is the proof that $c=\pm 1$ when $\phi_1(x)$ is algebraic.

† Koebe, loc. cit., or Ritt, these Transactions, vol. 23 (1922), p. 20.

9. We shall show that when the functions $\phi_i(x)$ are algebraic, c in (16) must be ± 1, except if β is 0. In the latter case, by (19), (16) becomes

$$\phi_i(x) = \phi_1\left(x + \frac{d_1}{c}\right), \quad \phi_j(x) = \phi_1\left(x + \frac{d_2}{c}\right), \quad \phi_k(x) = \phi_1\left(x + \frac{d_1 + d_2}{c}\right).$$

Thus, when the functions $\phi_i(x)$ are algebraic, ± 1 will always serve as values of c in (16).

We begin by observing that, when $\phi_1(x)$ is algebraic, every α_i in (3) must be real. Suppose that an α_i is not real. Then $\phi_1(x) = \phi_i(x - \alpha_i)$ assumes real values on the straight segment joining $x_{i-1} - \alpha_i$ to $x_i - \alpha_i$. This segment is parallel to the real axis. But $\phi_1(x)$ assumes real values in $(0, x_1)$. Hence it assumes conjugate values at conjugate points. Thus, given any point on the above mentioned segment, $\phi_1(x)$ assumes the same real value at that point and at its conjugate. The difference between the two conjugate points stays equal, as the first point varies over its segment, to twice the imaginary part of α_i. This constant difference must be a period of $\phi_1(x)$. But $\phi_1(x)$, being algebraic, has no periods. This proves that α_i is real.

From (16) and (3), we find that

$$\phi_k(x) = \phi_i\left(x + \frac{d_2}{c}\right) = \phi_1\left(x + \alpha_i + \frac{d_2}{c}\right).$$

Thus we may use $\alpha_i + d_2/c$ for α_k, so that d_2 is a real multiple of c.

Suppose that $c \neq 1$. Referring to the value of β, we see that, if c is not -1, β will not be real unless $\beta = 0$. Suppose that $c \neq -1$ and that $\beta \neq 0$. Let $\bar{\beta}$ represent the conjugate of β. Because $\phi_1(x)$ assumes conjugate values at conjugate points, there exists, corresponding to (19), a relation

$$\phi_1(x) = \phi_1[c(x - \bar{\beta}) + \bar{\beta}]. \tag{20}$$

Equating the second members of (19) and (20) we see that

$$(c - 1)(\beta - \bar{\beta}),$$

which does not vanish, is a period of $\phi_1(x)$. Thus, when c is not -1, we have $\beta = 0$. This shows that $\phi_i(x)$, $\phi_j(x)$ and $\phi_k(x)$ satisfy a relation (16) with $c = \pm 1$, and completes the proof of Theorem III.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

CONCERNING CONTINUA IN THE PLANE*

BY

GORDON T. WHYBURN

In this paper a study will be made of plane continua. Part I deals with continua which constitute the boundary of a connected domain and is concerned in particular with (1) properties of domains which are consequences of certain conditions imposed upon their boundaries, (2) properties of the boundaries of domains which are consequences of conditions imposed upon the domains, and (3) conditions under which the boundary of a domain is accessible from that domain. Part II is concerned with the cut points and end points of continua.

I wish to acknowledge my indebtedness to Professor R. L. Moore, to whom the success of this investigation should be largely attributed. Credit is due him for the suggestion of most of the problems treated in Part I; and it is his stimulating personality, constant encouragement, and many helpful suggestions and criticisms which have attracted my interest to this field of mathematics and have made possible the solution of the problems treated in this paper.

I. Domains and their boundaries

Definitions. A domain D is said to have property S† provided it is true that for any positive number ϵ, D may be expressed as the sum of a finite number of connected point sets each of diameter less than ϵ. A point set K will be said to be uniformly connected im kleinen with reference to every one of its bounded subsets provided it is true that if M is any bounded point set whatever and ϵ is any positive number, then there exists a positive number δ_ϵ such that every two points which are common to M and K and whose distance apart is less than δ_ϵ lie together in a connected subset of K of diameter less than ϵ. A boundary point P of a domain D is accessible from all sides from D‡ provided it is true that if A and B are any two points

---

* Various parts of this paper were presented to the Society under different titles, December 29, 1925, February 27, April 2, May 1, and June 12, 1926; it was received by the editors in its present form in June, 1926.

† Cf. R. L. Moore, *Concerning connectedness im kleinen and a related property*, Fundamenta Mathematicae, vol. 3 (1922), pp. 232–237.

‡ Cf. A. Schoenflies, *Die Entwickelung der Lehre von den Punktmannigfaltigkeiten*, zweiter Teil, Jahresbericht der Deutschen Mathematiker-Vereinigung, Ergänzungsbände, vol. 2 (1908), p. 215

of the boundary of D and AXB is an arc such that $AXB-(A+B)$ is a subset of D and such that AXB separates D into two domains D_1 and D_2, then P is accessible from every one of the domains D_1, D_2 to whose boundary it belongs. Two point sets will be said to be mutually separated if they are mutually exclusive and neither contains a limit point of the other. The point P of a continuum M will be called a cut point of M provided the set $M-P$ is not connected, i.e., is the sum of two mutually separated sets.

Notation. In this paper wherever a symbol X is used to denote a point set, the symbol $\overline{X}$ will be used to denote the set X plus all those points which are limit points of X. And wherever a symbol AB is used to designate a simple continuous arc, the symbol (AB) will be used to denote the point set $AB-(A+B)$.

THEOREM 1. *In order that a bounded domain D should have property S it is necessary and sufficient that every point of the boundary of D should be accessible from all sides from D.*

Proof. I shall first show that the condition is necessary. Suppose D is a domain having property S and P is a point of its boundary. Let A and B be two points of the boundary of D, and let AXB be an arc from A to B such that (AXB) is a subset of D, and such that AXB separates D into two domains D_1 and D_2. In an unpublished paper,* C. M. Cleveland has proved the following theorem. In order that a bounded domain D should have property S it is necessary and sufficient that (1) every maximal connected subset of the boundary of D should be a continuous curve, and (2) for any given positive number ϵ, there should be not more than a finite number of these continuous curves of diameter greater than ϵ. Now since D has property S, it follows that the boundary of D satisfies conditions (1) and (2) of Cleveland's theorem. And since the boundary of D satisfies these conditions, it can easily be shown by methods almost identical with those used by R. L. Moore to prove Theorem 4 of his paper *Concerning connectedness im kleinen and a related property*† that the boundary of D_1, and also that of D_2, must satisfy these conditions. Hence, it follows by Cleveland's theorem that each of the domains D_1 and D_2 must have property S. Now let R denote either one of the domains D_1, D_2 which has the point P in its boundary. It is sufficient, then, to show that P is accessible from R.

---

* Cf. an abstract of a paper by the author and C. M. Cleveland, Bulletin of the American Mathematical Society, vol. 32 (1926), p. 109.

† Loc. cit.

Let R be expressed as the sum of n_1 connected point sets $K_{11}, K_{12}, \cdots, K_{1n_1}$, all of diameter less than $1/3$. Let G_1 denote this collection of point sets. Let S_1 denote the collection of all those elements of G_1 which have P for a limit point, and let T_1 denote the sum of all the point sets of the collection S_1. There exists a circle C_1 having P as center and neither containing nor enclosing any point of $R-T_1$. Let X_1 denote a point common to T_1 and to the interior of C_1. Let I_1 denote the sum of all those elements of S_1 which can be joined to that element of S_1 which contains X_1 by a connected subset of R lying wholly within C_1. Every point of I_1 which is a limit point of $R-I_1$ lies within a circle c such that c plus its interior belongs to R and is of diameter less than $1/9$. Add to I_1 the interiors of all such circles (c), and let R_1 denote the domain thus obtained. Clearly R_1 is of diameter less than 1, and P is a boundary point of R_1. Now let R be expressed as the sum of n_2 connected point sets $K_{21}, K_{22}, K_{23}, \cdots, K_{2n_2}$, all of diameter less than $1/6$ and also less than the radius of C_1. Let G_2 denote this collection of point sets. And let T_2 and C_2 be point sets which, with respect to G_2, correspond to T_1 and C_1 selected previously with respect to G_1. Let X_2 be a point common to T_2, to the interior of C_2, and to I_1. Let I_2 denote the sum of all those elements of S_2 which can be joined to the element of S_2 which contains X_2 by a connected subset of R lying wholly within C_2. Clearly I_2 is a subset of I_1, and hence also of R_1. Every point of I_2 which is a limit point of $R-I_2$ lies within some circle c such that c plus its interior belongs to R and to R_1 and is of diameter less than $1/18$. Add to I_2 the interiors of all such circles (c), and let R_2 denote the domain thus obtained. Clearly R_2 is a subset of R_1, is of diameter less than $\frac{1}{2}$, and has the point P in its boundary. This process may be continued indefinitely, and thus we obtain a sequence of subdomains of R: $R_1, R_2, R_3, \cdots$, such that for every positive integer n, R_{n+1} has P in its boundary and is a subset of R_n, and such that the diameter of R_n approaches zero as a limit as n increases indefinitely.

Now let Q denote any point of R. For each positive integer n, let P_n denote a point of R_n. There exists an arc QP_1 lying in R, and for each n, there exists an arc P_nP_{n+1} lying in R_n. It is easy to see* that the point set $P+QP_1+P_1P_2+P_2P_3+\cdots$ is closed and that it contains as a subset an arc QP such that $QP-P$ is a subset of R. Hence, P is accessible from R, and since R is either one of the domains D_1, D_2 which has P in its boundary, it follows that P is accessible from all sides from D.

---

* Cf., for example, R. L. Moore, *On the foundations of plane analysis situs*, these Transactions, vol. 17 (1916), p. 155, paragraph beginning with line five.

The condition is also sufficient. Suppose D is a bounded domain such that its boundary, M, is accessible from all sides from D. Condition (I) will be said to exist if some maximal connected subset of the boundary of D fails to be a continuous curve; Condition (II) will be said to exist if it is true that for some positive number ϵ, there exist infinitely many maximal connected subsets of M of diameter greater than ϵ. Suppose Condition (II) exists, and let G denote the set of all those maximal connected subsets of M which are of diameter greater than ϵ. Since the sum of all the continua of G is bounded and G contains infinitely many elements, it follows* that there exists a continuum T of diameter at east as great as ϵ which is the sequential limiting set of some sequence $T_1, T_2, T_3, \cdots$ of elements of G. There exist two points E and F of T whose distance apart is $\geqq\epsilon$. Let C_1 be a circle with E as center and of radius $\frac{3}{4}\epsilon$. Let C_2 be a circle with E as center and of radius $\frac{1}{4}\epsilon$. Then since E is within C_2 and F is without C_1, there exists a positive integer d such that for every $n>d$, T_n contains a point x_n within C_2 and a point y_n without C_1. It follows from a theorem due to Janiszewski† that for every $n>d$, T_n contains a subcontinuum t_n which contains at least one point on each of the circles C_1 and C_2 and is a subset of the point set H consisting of the circles C_1 and C_2 together with all those points of the plane which lie between C_1 and C_2. For every positive integer i, let M_i denote t_{d+i}. The continuum T contains a subcontinuum M_∞ which has at least one point on each of the circles C_1 and C_2, is a subset of H, and is the sequential limiting set of the sequence $M_1, M_2, M_3, \cdots$. Now suppose Condition (I) exists. It follows directly from a theorem of R. L. Moore's‡ that there exist circles C_1 and C_2, and that M contains a countable infinity of continua M_∞, $M_1, M_2, M_3, \cdots$, having exactly the same properties as the point sets of the same notation whose existence was shown as a consequence of Condition (II). Hence, we see that the existence of either Condition (I) or Condition (II) leads to exactly the situation described above.

Let A (Fig. 1) denote a point common to M_∞ and C_2, and B a point common to M_∞ and C_1. Since M is accessible from D, it follows that there exists an arc AB such that (AB) is a subset of D. It can be shown that there exists a bounded complementary domain R of the point set $AB+M_\infty$ such

---

* See an abstract of a paper by R. G. Lubben, *Concerning limiting sets*, Bulletin of the American Mathematical Society, vol. 32 (1926), p. 14.

† *Sur les continus irréductibles entre deux points*, Journal de l'Ecole Polytechnique, (2), vol. 16 (1912).

‡ *A report on continuous curves from the viewpoint of analysis situs*, Bulletin of the American Mathematical Society, vol. 29 (1923), pp. 296–297.

that every point of the arc AB belongs to the boundary of R. It can be shown that the arc AB separates D into two domains D_1 and D_2 such that D_1 lies wholly within R, and D_2 lies wholly in K, the unbounded complementary domain of the boundary of R. Since no member of the sequence of continua $M_1, M_2, M_3, \cdots$ has a point in common with $AB+M_\infty$, it follows that for every positive integer i, M_i lies either wholly in R or wholly in K. Hence, either R or K must contain infinitely many members of the sequence $M_1, M_2, M_3, \cdots$.

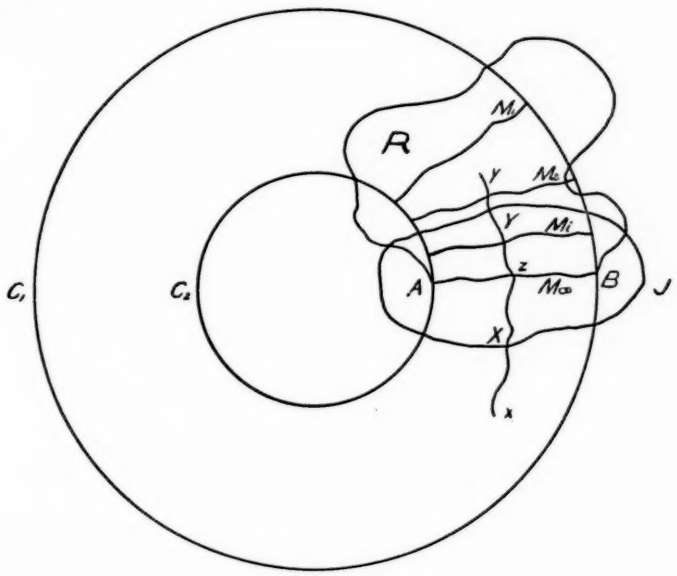

Fig. 1

Suppose R contains infinitely many members of this sequence. Then every point of M_∞ is a limit point of a set of points common to D and R. And since all such points belong to D_1, it follows that every point of M_∞ is a boundary point of D_1 and, by hypothesis, is therefore accessible from D_1. Let O be a point of K and let P be a point of M_∞ distinct from A and from B. Then since the arc AB does not of itself separate the plane, there exists an arc OP which contains no point of the arc AB. On OP, in the order from O to P, let z denote the first point belonging to M_∞. Then the point set $Oz-z$ is a subset of K. Now either (a) there exists a point x on the arc Oz such that the arc zx of Oz contains no point of M, or (b) z is a limit point of a set of points common to D and K. In case (b), since all points common to K and D

belong to D_2, then z is a boundary point of D_2 and is, therefore, accessible from D_2. Hence, if x is a point of D_2, there exists an arc xz such that $xz-z$ is a subset of D_2. Hence, in either case, (a) or (b), there exists an arc xz such that $xz-z$ is a subset of K and contains no point whatever of M. It was shown above that z is accessible from D_1. Hence, if y is a point of D_1, there exists an arc yz such that $yz-z$ is a subset of D_1. The two arcs xz and yz can have in common only the point z.

Let I denote the point set consisting of M_∞ plus all of its bounded complementary domains. Let L denote the closed point set $I+M_1+M_2+M_3+\cdots$. It can easily be shown that neither of the points x and y belongs to I. Now I does not separate the plane, and it is a maximal connected subset of the closed set L. By a theorem of R. L. Moore's* it follows that there exists a simple closed curve J such that J encloses I and contains no point of L, and every point on or within J is at a distance from some point of I less than the minimum distance from x to I and also less than the minimum distance from y to I. Hence both x and y are without J. On the arc zx, in the order from z to x, and on zy, in the order from z to y, let X and Y respectively denote the first points belonging to J. Denote the two arcs of J from X to Y by XTY and XSY respectively, and let R_1 and R_2 denote the interiors of the closed curves $XzYTX$ and $XzYSX$ respectively. On the arc XzY there exist points E, U, H, and G in the order X, E, U, z, H, G, Y such that within some circle which has z as center and which neither contains nor encloses any point of the arc AB there exist arcs EFG and UCH such that (EFG) and (UCH) are subsets of R_1 and R_2 respectively. Since E and U lie in K, and H and G lie in R, it follows that both (EFG) and (UCH) must contain a point in common with M_∞. But (EFG) belongs to R_1, and (UCH) belongs to R_2. Hence R_1 contains a point u of M_∞, and R_2 contains a point v of M_∞. Let C_u and C_v be circles having u and v respectively as centers and such that C_u belongs to R_1 and C_v belongs to R_2. Now since J encloses M_∞ and contains no point of L, there exists a positive number d_1 such that for every integer $n>d_1$, M_n lies wholly within J. There exists a positive number d_2 such that for every integer $n>d_2$, M_n has a point within C_u and also a point within C_v. Let i be an integer which is greater than each of these two numbers d_1 and d_2. Then M_i lies wholly within J and contains at least one point in each of the domains R_1 and R_2. Hence, since it is connected, M_i must contain a point p of the arc XzY. Since M_i has no point in common with M_∞, therefore $p\neq z$. Hence p must belong either to (zX) or to (zY). But p

* *Concerning the separation of point sets by curves*, Proceedings of the National Academy of Sciences, vol. 11 (1925), p. 469.

belongs to M, and neither (zX) nor (zY) contains any point whatever of M. Thus in case R contains infinitely many members of the sequence $M_1, M_2, M_3, \cdots$, the supposition that either Condition (I) or Condition (II) exists leads to a contradiction. In an entirely similar manner the same supposition leads to a contradiction in case K contains infinitely many members of this sequence.

Since neither Condition (I) nor Condition (II) can exist, then (1) every maximal connected subset of the boundary of D is a continuous curve, and (2) for every positive number ϵ, there are not more than a finite number of these maximal connected subsets of diameter greater than ϵ. Since D is bounded, it follows from Cleveland's theorem quoted above that D has property S.

THEOREM 2. *If the domain D is uniformly connected im kleinen with reference to every one of its bounded subsets, then every point of the boundary of D is accessible from D.*

Proof. Let P denote a point of the boundary of D. Let C_1 be a circle having P as center and of diameter less than 1. For every point x common to D and the interior of C_1, let K_x denote the greatest connected point set which contains x and is common to D and the interior of C_1. Let G_1 denote the collection of all such sets (K_x). Since D is uniformly connected im kleinen with respect to every one of its bounded subsets, it follows that if C is any circle concentric with and within C_1, then there are not more than a finite number of elements of G_1 which have points on or within C and have a limit point on C_1. Hence there exists a circle k_1, concentric with and within C_1, such that k_1 neither contains nor encloses any point of any element of G_1 which does not have P for a limit point. Let S_1 denote the finite collection of all those elements of G_1 which have points on or within k_1, and let T_1 denote the sum of all the point sets of this collection. Let X_1 be a point common to T_1 and the interior of k_1. Let D_1 denote the element of S_1 which contains X_1. Clearly D_1 is a domain which (1) is a subset of D and of the interior of C_1, (2) has P in its boundary, and (3) contains every point common to D and the interior of C_1 which can be joined to X_1 by an arc which is also a subset of D and of the interior of C_1. Now let C_2 denote a circle which is concentric with C_1 and of diameter less than $\frac{1}{2}$ and also less than the radius of k_1. Let G_2, k_2, S_2, T_2, X_2, and D_2 be collections and sets which with respect to C_2 correspond to G_1, k_1, S_1, T_1, X_1, and D_1 selected above with respect to C_1, with the additional condition that X_2 shall belong to D_1. Then D_2 is a domain which (1) is a subset of D, of D_1, and of the interior of C_2, (2) has P in its boundary, and (3) contains every point common to D and the interior

of C_2 which can be joined to X_2 by an arc which is also a subset of D and of the interior of C_2. This process may be continued indefinitely, and thus we obtain a sequence of subdomains of D: $D_1, D_2, D_3, \cdots$, such that for every positive integer n, D_{n+1} has P in its boundary and is a subset of D_n, and such that the diameter of D_n approaches zero as a limit as n increases indefinitely. By an argument which is identical with the third paragraph of the proof of Theorem 1, it follows that if A is any point of D, then there exists an arc AP such that $AP-P$ is a subset of D. Hence, every point of the boundary of D is accessible from D.

THEOREM 3. *In order that a domain D should be uniformly connected im kleinen with reference to every one of its bounded subsets it is necessary and sufficient that* (1) *every maximal connected subset of the boundary of D should be either a point, a simple closed curve, or an open curve, and* (2) *if ϵ is any positive number and J is any simple closed curve, there should be not more than a finite number of maximal connected subsets of the boundary of D which have points within J and are of diameter greater than ϵ.*

Proof. I shall show that the conditions are necessary. This may be done with the use of methods only slightly different from those used by R. L. Moore in his paper *A characterization of Jordan regions by properties having no reference to their boundaries*† to prove the proposition that every bounded, simply connected, and uniformly connected im kleinen domain is bounded by a simple closed curve. I shall simply indicate the modifications necessary in his argument to establish Theorem 3.

Suppose the domain D is uniformly connected im kleinen with reference to every one of its bounded subsets. Then by an argument almost identical with that used by Moore to show that the boundary of his domain in the above mentioned proposition is a continuous curve, it follows that every maximal connected subset of the boundary of D is a continuous curve, and that if J is any simple closed curve and ϵ is any positive number, then there are not more than a finite number of these maximal connected subsets of the boundary of D which have points within J and are of diameter greater than ϵ. Now let M denote any definite maximal connected subset of the boundary of D which consists of more than one point. I shall show that M must be either a simple closed curve or an open curve. Let the points $P_1, P_2, P_3, \cdots$, the arcs $A_1B_1, A_2B_2, A_3B_3, \cdots$, and the point set N^* be selected and defined with respect to M exactly as was done by Moore in the paragraph beginning

† Proceedings of the National Academy of Sciences, vol. 4 (1918), pp. 364–70.

near the bottom of page 366 of his paper. I shall now show that M is neither a simple continuous arc nor a ray of an open curve. Suppose the contrary is true. Then if M is an arc, let A and B denote its end points, and if M is a ray, let A denote its end point. Let X be a point of M which is distinct from A and from B, and let C be a circle with X as center and not enclosing or containing either A or B. Within C and on M there exist points E, U, W, and G in the order A, E, U, X, W, G, and within C there exist arcs EFG and UVW having only their end points in common with M and such that if R_1 and R_2 denote the interiors of the closed curves $EFGWXUE$ and $UVWXU$ respectively, then R_1 and R_2 are mutually exclusive domains each of which lies wholly within C. Since under this supposition, M can contain no simple closed curve, it readily follows that X must be a limit point of a set of points K_1 common to D and R_1 and also of a set K_2 common to D and R_2. But clearly this is impossible, since D is uniformly connected im kleinen with reference to every one of its bounded subsets. It follows, then, that M is neither an arc nor a ray of an open curve.

Now suppose M is bounded. In this case, since M cannot be an arc, it follows by exactly the same argument given by Moore in the first paragraph of page 369 of his paper that M is a simple closed curve. Suppose M is unbounded. Since M cannot be a ray of an open curve, it readily follows that both of the sequences of points $A_1, A_2, A_3, \cdots$ and $B_1, B_2, B_3, \cdots$ must be infinite and that neither of these sequences can have a limit point. It follows that N^* is a closed point set which is identical with M and which, evidently, must be an open curve. Hence the conditions are necessary.

C. M. Cleveland† has proved that the conditions are sufficient.

THEOREM 4. *If K denotes the set of all the cut points of the boundary M of a complementary domain D of a continuous curve, then $D+K$ is uniformly connected im kleinen.*

Proof. By a theorem due to Miss Torhorst,‡ M is a continuous curve. Suppose $D+K$ is not uniformly connected im kleinen. Then, for some positive number ϵ, D contains two infinite sequences of points, $X_1, X_2, X_3, \cdots$ and $Y_1, Y_2, Y_3, \cdots$, such that (1) for every positive integer n, the distance from X_n to Y_n is less than $1/n$, (2) for no integer n is it true that X_n and Y_n lie together in some connected subset of $D+K$ of diameter less than ϵ, and (3) there exists in M a point P which is the sequential limiting

† In unpublished work.

‡ *Ueber den Rand der einfach zusammenhängenden ebenen Gebiete*, Mathematische Zeitschrift, vol. 9 (1921), p. 64 (73).

set of each of these two sequences of points. Let C be a circle having P as center and of diameter $\epsilon/2$. Let us first suppose that D is bounded. It follows by a theorem of R. L. Moore's* that D has property S. Hence, D is expressible as the sum of a finite number of connected point sets K_1, K_2, K_3, $\cdots$, K_n, all of diameter less than $\epsilon/5$. Let K_{m_1}, K_{m_2}, K_{m_3}, $\cdots$, K_{m_m} denote those sets of this sequence which have P as a limit point. Clearly $K_{m_1}+K_{m_2}+\cdots+K_{m_m}$ is a subset of the interior of C. Since P is not a limit point of $D-(K_{m_1}+K_{m_2}+\cdots+K_{m_m})$, there exists a positive integer i such that both X_i and Y_i belong to $K_{m_1}+K_{m_2}+\cdots+K_{m_m}$. Let N_x and N_y denote sets of this sequence which contain X_i and Y_i respectively. Let R_x and R_y denote the maximal connected subsets of D which contain N_x and N_y respectively and lie within C. Clearly, the domains R_x and R_y can have no points in common. The point P belongs to the boundary of each of these domains, and by the method used in the proof of Theorem 1, it can be shown that P is accessible from each of them. Hence, there exist arcs X_iP and Y_iP such that X_iP-P and Y_iP-P are subsets of R_x and R_y respectively. There exists an arc t from X_i to Y_i which is a subset of D. The point set $t+X_iP+Y_iP$ contains a simple closed curve J which contains P and lies, except for the point P, wholly in D. Let I and E denote the interior and exterior respectively of J. If either I or E, say I, contains no point of M, then since D contains points of I, it follows that I is a subset of D, and clearly in this case X_i and Y_i can be joined by a connected subset of D of diameter less than ϵ, contrary to supposition. And if both I and E contain points of M, then clearly P is a cut point of M and belongs to K. And in this case R_x+R_y+P is a connected subset of $D+K$ which contains both X_i and Y_i and is of diameter less than ϵ, contrary to supposition. Thus, in any case, the supposition that $D+K$ is not uniformly connected im kleinen leads to a contradiction. The case where D is unbounded may be treated in a slightly modified manner.

THEOREM 5. *In order that the simply connected bounded domain D should become uniformly connected im kleinen upon the addition of a single point O of its boundary B it is necessary and sufficient that* (1) *if K be any maximal connected subset of $B-O$, then $K+O$ is a simple closed curve, and* (2) *there be not more than a finite number of these curves of B of diameter greater than any preassigned positive number.*

Proof. The conditions are necessary. Suppose D is a bounded domain with connected boundary B, and O is a point of B such that $D+O$ is uniformly

* *Concerning connectedness im kleinen and a related property*, loc. cit., Theorem 4.

connected im kleinen. Then B is a continuous curve. For suppose it is not. Then B contains a point P which is distinct from O and at which B is not connected im kleinen. Then by an argument identical with that used by R. L. Moore in his paper *A characterization of Jordan regions by properties having no reference to their boundaries*,† in the paragraph beginning at the bottom of page 365, with the additional condition that the circle K used in his argument be taken of radius less than $\frac{1}{2}$ the distance between O and P, it can be shown that this supposition leads to a contradiction. Hence, B is a continuous curve.

Let K denote a maximal connected subset of $B-O$. Then since $B-K$ is closed, it follows that K is connected im kleinen. Now let an inversion of the plane be performed about some circle which has O as center. Since $K+O$ is closed and connected, it follows that K^*, the image of K, is unbounded, closed, connected, and connected im kleinen. Since the inversion does not act upon the point O, and since $D+O$ is uniformly connected im kleinen, it can readily be shown that D^*, the image of D, is uniformly connected im kleinen with reference to every one of its bounded subsets. Therefore, by Theorem 3, it follows that K^* is an open curve, and hence, that $K+O$ is a simple closed curve. Therefore, condition (1) is necessary. Now since every maximal connected subset K of $B-O$ is a simple closed curve minus one point, every such set K contains an arc of diameter greater than $\frac{1}{2}$ the diameter of K. By a theorem of R. L. Wilder's,‡ B cannot contain, for any given positive number ϵ, more than a finite number of mutually exclusive arcs all of diameter greater than ϵ. In view of this result, it follows that for any positive number ϵ, $B-O$ cannot contain an infinite number of maximal connected subsets each of diameter greater than ϵ. Hence, condition (2) is necessary.

The conditions are also sufficient. Suppose D is a bounded domain with connected boundary B which satisfies conditions (1) and (2) in the statement of this theorem. Clearly, B must be a continuous curve. Unless the point O is a cut point of B, then B is a simple closed curve and D is its interior. In this case D itself is uniformly connected im kleinen. Hence, unless this theorem is true, O must be a cut point of B. No other point is a cut point of B. For let P denote any other point of B. Let K denote the maximal connected subset of $B-O$ which contains P, and let J denote the point set $K+O$. By hypothesis, J is a simple closed curve. Hence, $J-P$ is connected. But $B-K$ is connected, and since the connected sets $J-P$ and

† Loc. cit.

‡ *Concerning continuous curves*, Fundamenta Mathematicae, vol. 7 (1925), p. 358, Theorem 6

$B-K$ have the point O in common, their sum S is connected. But $S=B-P$. Hence, P is not a cut point of B. It follows that O is the only cut point of B, and therefore, by Theorem 4, $D+O$ is uniformly connected im kleinen.

THEOREM 6. *In order that a continuous curve M should be the boundary of a connected domain it is necessary and sufficient that if J denotes any simple closed curve of M, then* (1) *M is a subset either of $J+I$ or of $J+E$, where I and E denote the interior and exterior respectively of J, and* (2) *if A and B are any two points of J, then $M-(A+B)$ is not connected.*

Proof. The conditions are necessary. That condition (1) is necessary is evident. Now let A and B denote any two points of J, where J is any simple closed curve contained in the boundary M of a complementary domain D of a continuous curve. Since A and B are accessible from D, it readily follows that there exists an arc AXB such that (AXB) is a subset of D. Now $M+D$ lies wholly either in J plus its interior I, or in J plus its exterior E, suppose in $J+I$. Then there exists an arc AYB such that (AYB) is a subset of E. Let t and t' denote the two arcs of J from A to B. Then the simple closed curve $AXBYA$ encloses one of these arcs minus $A+B$, say $t-(A+B)$, and neither contains nor encloses any point of $t'-(A+B)$. Since M has in common with the curve $AXBYA$ only the points A and B, it follows that $M-(A+B)$ is not connected. Hence the conditions are necessary.

The conditions are also sufficient. Let M denote a continuous curve which satisfies conditions (1) and (2) of this theorem. Let K denote the unbounded complementary domain of M, and let N denote its boundary. Now N contains a simple closed curve J, or otherwise M† is the boundary of K and the theorem is true. By hypothesis, M is a subset either of $J+I$ or of $J+E$, where I and E denote the interior and exterior respectively of J.

Case I. Suppose M is a subset of $J+E$. I shall show that in this case $N\equiv M$, i.e., that M is the boundary of K. Suppose M contains a point P which does not belong to N. Then let R denote the complementary domain of N which contains P and let C denote its boundary. By a theorem of R. L. Moore's‡ it follows that C is a simple closed curve. Since R is bounded, C encloses P; and P belongs to E, the exterior of J. Hence, J contains a point Q which does not belong to C. The curve C does not enclose Q, for Q is a boundary point of K, the unbounded complementary domain of M.

† R. L. Moore, *Concerning continuous curves in the plane*, Mathematische Zeitschrift, vol. 15 (1922), pp.254–260, Theorem 5.

‡ Loc. cit., Theorem 4.

Hence Q lies in the exterior of C. But C encloses P and, by hypothesis, M is a subset either of C plus its interior or of C plus its exterior. Thus the supposition that $N \neq M$ leads to a contradiction. Hence M is the boundary of the connected domain K.

Case II. Suppose M is a subset of $J+1$. With the aid of hypothesis (2) it is shown that there exists a point O which does not belong to M and which is within J but is not within any other simple closed curve belonging to M. Let C be a circle having O as center and not enclosing or containing any point of M. Let an inversion of the plane be performed about C. If X is a point set, let X' denote the image of X under this inversion. Now M' is a subset of $J'+I'$, and I' is the exterior of J'. Let K' denote the unbounded complementary domain of M', and let N' denote its boundary. Then N' contains J', and by an argument identical with that used in Case I it can be shown that M' is the boundary of the connected domain K'. Hence, it follows that M is the boundary of the connected domain K, where K is the point set of which K' is the image under this inversion of the plane.

THEOREM 7. *If the point P of a continuous curve M belongs to the boundary of no complementary domain of M, then for every positive number ϵ, M contains a simple closed curve which encloses P and is of diameter less than ϵ.*

Proof. Let P denote a point of a continuous curve M which belongs to the boundary of no complementary domain of M, and let ϵ denote any positive number. Let C be a circle having P as center and of diameter less than $\epsilon/2$, and such that the exterior of C contains at least one point of M. Let N denote the maximal connected subset of M which contains P and is contained in C plus its interior. By a theorem of H. M. Gehman's,* N is a continuous curve. The curve N contains a point A which belongs to C. Any arc whatever from A to P must contain at least one point of N which is distinct from A and from P. For suppose there exists an arc t from A to P which has only the points A and P in common with N. Since M is connected im kleinen, it readily follows that P is not a limit point of $M-N$. Hence, there exists a point X on t such that the arc PX of t has only the point P in common with M. Hence, the connected set $PX-P$ belongs to some complementary domain of M, and P must be a boundary point of that domain. But P is not a boundary point of any complementary domain of M. It follows, then, that every arc from A to P contains a point of N which is

* *Concerning the subsets of a plane continuous curve*, Annals of Mathematics, (2), vol. 27 (1925), pp. 29–46, Theorem 4, Lemma B.

distinct from A and from P. By a theorem proved by C. M. Cleveland,* it follows that N contains a simple closed curve J which encloses either A or P. The curve J cannot enclose A, because A belongs to C, and J is a subset of C plus its interior. Hence J must enclose P. Since it is contained in C plus its interior, J is of diameter less than ϵ.

II. Cut points and end points

In this section, I shall make a study of the properties of the cut points and end points of a given plane continuum. More particularly, I shall study the connected subsets of the set of all the cut points and end points of a continuum, and I shall establish some very fundamental properties of such sets, both internal properties and properties relative to the remainder of the continuum.

Definitions. The term *cut point* will be used as defined in Part I. The term *end point*, as applied to a continuous curve, will be used in the sense as defined by R. L. Wilder,† i.e., a point P of a continuous curve M will be called an end point of M provided it is true that if t is any arc of M having P as one of its extremities, then $M-(t-P)$ contains no connected subset which contains P. As applied to continua in general, I shall define the term *end point* as follows. The point P of a continuum M will be called an end point of M provided it is true that if N is any subcontinuum of M which contains P, then P is not a limit point of any connected subset of $M-N$. It is obvious that this definition will allow as many, if not more, points of a continuum to be end points as would the following extension of Wilder's definition: the point P of a continuum M is said to be an end point of M provided it is true that if H is any subcontinuum of M which contains P, then P belongs to no connected subset of $M-(H-P)$. The term *acyclic continuous curve* will be used, after Gehman, to designate a continuous curve which contains no simple closed curve.

R. L. Moore has shown‡ that no subcontinuum K of a given continuum M can contain an uncountable set of points each of which is a cut point of M but not of K. It follows from this theorem that no simple closed curve K can contain more than a countable number of cut points of any continuum

---

* This theorem is to the effect that if A and P are distinct points of a continuous curve N and every arc from A to P contains at least one point of N distinct from A and from P, then N contains a simple closed curve which separates A from P. Cf. an abstract of a paper by C. M. Cleveland, Bulletin of the American Mathematical Society, vol. 32 (1926), p. 311.

† *Concerning continuous curves*, Fundamenta Mathematicae, vol. 7 (1925), p. 358.

‡ *Concerning the cut points of continuous curves and of other closed and connected point sets*, Proceedings of the National Academy of Sciences, vol. 9 (1923), pp. 101–106.

which contains K. Extensive use will be made of these results in the proofs given in this section.

THEOREM 8. *If H is any connected subset of a continuum M, then not more than a countable number of points of $\overline{H}-H$ are cut points of M.*

Proof. Let T denote the set of all those points of $\overline{H}-H$ which are cut points of M. Clearly no point of T is a cut point of $\overline{H}$. Hence, by R. L. Moore's theorem quoted above, it follows that T must be countable.

THEOREM 9. *If K denotes the set of all the cut points and H the set of all the end points of a continuum M, and if T is any countable subset of M, then every bounded, closed, and connected subset of $K+H+T$ is an acyclic continuous curve.*

Proof. Let N denote any bounded continuum which is a subset of $K+H+T$. I shall first show that N is a continuous curve. Suppose N is not a continuous curve. Then by R. L. Moore and R. L. Wilder's* characterization of continua which are not continuous curves it follows that there exist two concentric circles k_1 and k_2 and that N contains a countable infinity of mutually exclusive continua N_∞, N_1, N_2, N_3, $\cdots$, such that (1) each of these continua contains at least one point on each of the circles k_1 and k_2, (2) the set N_∞ is the sequential limiting set of the sequence of sets N_1, N_2, N_3, $\cdots$, and (3) there exists a connected subset L of N which contains all of the sets of the sequence N_1, N_2, N_3, $\cdots$, but which contains no point whatever of the set N_∞. Now clearly $\overline{L}-L$ contains the continuum N_∞. Hence, by Theorem 8, N_∞ can contain not more than a countable number of points of K. And since every point of N_∞ is a limit point of L, a connected subset of $M-N_\infty$, it follows that no point whatever of N_∞ can belong to H. Hence, since T is countable and N_∞ is a subset of $K+H+T$, it follows that N_∞ is countable. But this is absurd. Thus the supposition that N is not a continuous curve leads to a contradiction.

Now suppose N contains a simple closed curve J. Then clearly no point of J can belong to H. And by R. L. Moore's theorem, only a countable number of points of J can belong to K. Hence, since T is countable, J must be countable. But this is impossible. It follows, then, that N is an acyclic continuous curve.

---

* See R. L. Moore, Proceedings of the National Academy of Sciences, vol. 9 (1923), p. 103, and R. L. Wilder, Fundamenta Mathematicae, vol. 7 (1925), p. 371.

THEOREM 10. *If K is any closed and connected subset of the set of all the cut points of a bounded continuum M, and H is any connected subset of $M-K$, then $\overline{H}$ and K have at most one point in common. And if H is a maximal connected subset of $M-K$, then $\overline{H}$ and K have exactly one point in common.*

Proof. Suppose, on the contrary, that for some closed and connected subset K of the set of all the cut points of a bounded continuum M, $M-K$ contains a connected set H such that $\overline{H}$ and K have two points A and B in common. Now since, by Theorem 9, K is a continuous curve, it follows that K contains an arc t from A to B. By Theorem 8, t contains only a countable number of points of $\overline{H}$. Hence, t contains an interior point O which does not belong to $\overline{H}$. Let C denote a circle enclosing O and not enclosing or containing any point of $\overline{H}$. Within C there exist points E, G, U, and W on t in the order A, E, U, O, W, G, B, and arcs EFG and UVW having only their end points in common with t and such that if D_1 and D_2 denote the interiors of the closed curves $EFGWOUE$ and $UVWOU$ respectively, then D_1 and D_2 are mutually exclusive domains each of which lies within C. Let N denote the continuum $\overline{H}+t$. Let X and Y denote points of D_1 and D_2 respectively, and let Z denote a point belonging to the unbounded complementary domain of M. It is readily seen that every arc from X to Y contains at least one point of N, and that not both X and Y can be joined to Z by an arc which contains no point of N. Let v denote one of the points X, Y which cannot be so joined to Z, and let u denote the other one of the points X, Y. Let R_v denote that complementary domain of N which contains v. Let β denote the boundary of R_v. Then let R_u denote that complementary domain of β which contains u, and let α denote its boundary. R. L. Moore has shown* that under these conditions α contains no cut point of itself. But since R_v contains that one of the domains D_1 and D_2 which contains v and R_u contains the one which contains u, it readily follows that α contains the arc WOU of t. But WOU belongs to K, and every point of K is a cut point of M. Hence α contains an uncountable set of points each of which is a cut point of M but not of α, and since α is a continuum, this conclusion is contradictory to R. L. Moore's theorem quoted above. Thus the supposition that $\overline{H}$ and K have more than one point in common leads to a contradiction.

Now if H is any *maximal* connected subset of $M-K$, it is clear that K must contain at least one limit point of H. And in view of the above argument it follows that $\overline{H}$ and K must have exactly one point in common.

---

* *Concerning the sum of a countable number of mutually exclusive continua in the plane*, Fundamenta Mathematicae, vol. 6 (1924), p. 190.

THEOREM 11. *If L denotes the set of all the cut points of a bounded continuum M, T is any countable subset of M, K is any closed and connected subset of $L+T$, and H is any connected subset of $M-K$, then K contains at most one limit point of H.*

Theorem 11 may be proved by an argument only slightly different from that given in the proof of Theorem 10.

THEOREM 12. *In order that the point P of a continuous curve M should be an end point of M it is necessary and sufficient that no arc of M should have P as one of its interior points.*

Proof. The condition is sufficient. Let P denote any point of M which is not an end point of M. I will show that every such point is an interior point of some arc of M. From the definition of an end point it follows that M contains some arc t having extremities at P and some other point A of M and such that $M-(t-P)$ contains a connected set N which contains P. Let X denote a point of N which is distinct from P. Let K denote the maximal connected subset of $M-t$ which contains X. I will show that P is a limit point of K. Suppose, on the contrary, that P is not a limit point of K. Let T denote the set of points common to N and K. Since M is connected im kleinen at every one of its points and t is closed, it readily follows (1) that $N-T$ contains no limit point of T, and (2) that T contains no limit point of $N-T$. Hence, N is expressible as the sum of two mutually separated point sets T and $N-T$. But this is impossible, because N is connected. It follows, then, that P is a limit point of K. Now K is a domain with respect to M,† for t is a closed set of points. And the boundary U of K with respect to M is a subset of t. Hence U contains no continuum of condensation. By a theorem of R. L. Wilder's‡ it follows that every point of U is accessible in M from K. I have just shown that P belongs to U. Hence, if B denotes a point of K, there exists an arc BP such that $BP-P$ is a subset of K. The arcs t and BP have in common only the point P. Hence, their sum, $t+BP$, is an arc APB from A to B which contains P as an interior point. I have shown, then, that every point of M which is not an end point of M is an interior point of some arc of M. It follows that every point of M which is not an interior point of any arc of M is an end point of M.

The condition is also necessary.§ For suppose some arc APB of M contains as an interior point the point P which is an end point of M. Clearly

† Cf. R. L. Wilder, *Concerning continuous curves*, loc. cit., Section I.

‡ Loc. cit., Theorem 1.

§ R. L. Wilder gives a proof of this part of the theorem for the special case of an acyclic continuous curve. His method of argument, however, is not applicable to the general case here treated.

this is impossible, because the arc PB of APB is a connected subset of $M-(AP-P)$ which contains P.

I will remark that Theorem 12 shows the equivalence of Wilder's definition of an end point of a continuous curve and the following one: the point P of a continuous curve M is said to be an end point of M provided it is true that if t is any arc of M having P as one of its extremities, then P is not a limit point of any connected subset of $M-t$. This latter definition for the case of a continuous curve is analogous to the one I have given above for continua in general.

THEOREM 13. *If K is a connected subset of the set of all the cut points of a continuous curve M, then in order that $\overline{K}$ should be an acyclic continuous curve it is necessary and sufficient that every point of $\overline{K}$ should be either a cut point or an end point of M.*

Proof. That the condition is sufficient is a corollary to Theorem 9. I will show that it is necessary. Suppose K is a connected set of cut points of a continuous curve M such that $\overline{K}$ is an acyclic continuous curve. Let P denote a point of $\overline{K}$ which is not an end point of M. I will show that P is a cut point of M. Let U denote a point of K which is distinct from P. Then $\overline{K}$ contains an arc t from U to P. Every point of t, except possibly the point P, is a cut point of M. For suppose t contains an interior point O which is not a cut point of M. Then O does not belong to K. Since, by a theorem of R. L. Wilder's,* every connected subset of an acyclic continuous curve is arcwise connected, it follows that $K+P$ contains an arc t_0 from U to P which does not contain O. Then the sum of the arcs t_0+t contains a simple closed curve, contrary to the hypothesis that $\overline{K}$ is acyclic. Hence, every point of t, except possibly the point P, is a cut point of M. Now since P is not an end point of M, it follows by Theorem 12 that M contains an arc APB having P as one of its interior points. Not both of the arcs AP and PB of APB can contain an interval in common with t which contains P, because P is an end point of t. Suppose AP has no interval in common with t which contains P. Then AP and t have in common only the point P. For suppose they have in common a point V which is distinct from P. The interval VP of AP contains a point Q which does not belong to t. In the order from Q to P and from Q to A respectively on AP, let X and Y denote the first points belonging to t. The simple closed curve formed by the arc XY of t plus the arc XQY of AP contains a segment XY every point of which is a cut point

* *Concerning continuous curves*, loc. cit., Theorem 20.

of M. Clearly this is impossible. Hence, it follows that AP and t have in common only the point P.

Now suppose, contrary to this theorem, that P is not a cut point of M. Then by a theorem of R. L. Moore's,* $M-P$ contains an arc b from U to A. The sum of the arcs $AP+t+b$ contains a simple closed curve J which contains a segment of t, every point of which is a cut point of M. This is absurd, and thus the supposition that P is not a cut point of M leads to a contradiction. It follows, then, that every point of $\overline{K}$ is either a cut point or an end point of M.

THEOREM 14. *If K denotes the set of all the cut points of a continuous curve M, then for every positive number ϵ, K contains not more than a finite number of mutually exclusive continua each of diameter greater than ϵ.*

Proof. Suppose Theorem 14 is not true. Then there exists a positive number ϵ such that K contains infinitely many mutually exclusive continua each of diameter greater than ϵ. Since by Theorem 9, every closed and connected subset of K is a continuous curve, it follows that K contains infinitely many mutually exclusive arcs each of diameter greater than ϵ. Let $t_1, t_2, t_3, \cdots$ denote some sequence of these arcs which has a sequential limiting set t. It is evident that t contains two points A and B whose distance apart is $\geqq\epsilon$. Now since M is uniformly connected im kleinen, there exists a positive number δ_ϵ such that every two points of M whose distance apart is less than δ_ϵ are end points of an arc of M of diameter less than $\frac{1}{4}\epsilon$. There exists a positive number d such that for every integer $n>d$, t_n contains a point X_n and a point Y_n whose distances from A and B respectively are less than $\frac{1}{2}\delta_\epsilon$. Let i and j denote two integers greater than d. Then X_i and X_j and also Y_i and Y_j can be joined by an arc of M of diameter less than $\frac{1}{4}\epsilon$. Let X_iX_j and Y_iY_j denote these two arcs. It is readily seen that the sum of the arcs $t_i+t_j+X_iX_j+Y_iY_j$ contains a simple closed curve J which contains an interval of the arc t_i. But every point of t_i is a cut point of M. Thus the supposition that Theorem 14 is false leads to a contradiction.

THEOREM 15. *If K is any closed and connected subset of the set of all the cut points of a continuous curve M, then for every positive number ϵ, $M-K$ contains not more than a finite number of maximal connected subsets of diameter greater than ϵ.*

Proof. Suppose Theorem 15 is not true. Then there exists a positive number ϵ such that $M-K$ contains an infinite collection G of maximal

* *Concerning continuous curves in the plane*, Mathematische Zeitschrift, vol. 15 (1922), p. 255.

connected subsets each of diameter greater than ϵ. By Theorem 10, K contains exactly one limit point of each set of the collection G. For each set g of G let X denote the limit point of g which belongs to K, and let H denote the set of all such points (X) thus defined. Now if H contains infinitely many distinct points, then K contains a point A which is a limit point of H. And if H contains only a finite number of points, then H contains a point A which is a limit point of each of an infinite number of distinct sets of the collection G. Let us first suppose that A is a limit point of H. Then H contains an infinite sequence of points $X_1, X_2, X_3, \cdots$ which has A as its sequential limit point. For every positive integer n, let G_n denote an element of G which has X_n as a limit point. The sequence of sets $G_1, G_2, G_3, \cdots$ have a sequential limiting set L which contains A. And since every element of G is of diameter greater than ϵ, it follows that L contains a point B whose distance from A is $\geqq\epsilon/3$. Now since M is connected im kleinen, it can easily be shown that B must belong to K. Let C_1 and C_2 be circles having A and B respectively as centers and each of diameter less than $\epsilon/10$. The sequence of points $X_1, X_2, X_3, \cdots$ contains an infinite subsequence $X_{n_1}, X_{n_2}, X_{n_3}, \cdots$ every point of which is within C_1. There exists a circle C_b having B as center and such that every point of M which is enclosed by C_b can be joined to B by an arc common to M and to the interior of C_2. There exists an integer i such that G_{n_i} contains a point V within C_b. Hence, M contains an arc t from V to B which lies within C_2. On t, in the order from V to B, let E denote the first point belonging to K. Then E is a limit point of G_{n_i}. But X_{n_i} is also a limit point of G_{n_i}, and X_{n_i} lies within C_1. Hence, K contains two distinct limit points of G_{n_i}. But this is contrary to Theorem 10. A similar conclusion is reached when it is assumed that A is a limit point of each of an infinite number of distinct elements of G. Thus the supposition that Theorem 15 is false leads to a contradiction.

THEOREM 16. *If the bounded continuum M has the property that every connected subset of M is arcwise connected, and K is any maximal connected subset of the set of all the cut points of M, and H denotes the set of all those limit points of K which K does not contain, then every point of H is an end point of M.*

Proof. Suppose, on the contrary, that H contains a point P which is not an end point of M. Now by a theorem of R. L. Wilder's,* M is a continuous curve. Hence by Theorem 12 M contains an arc APB having P as one of its interior points. Let U denote a point of K. By hypothesis $K+P$ contains an arc t from U to P. Now P is not a cut point of M, for otherwise it would

* *Concerning continuous curves*, loc. cit., Theorem 18.

belong to K. In view of this fact, it follows by an argument almost identical with the latter part of the proof of Theorem 13, beginning with the fifteenth sentence, that this situation leads to an absurdity. Hence, every point of H is an end point of M.

THEOREM 17. *Under the same hypothesis as in Theorem* 16, $K+H$ *is an acyclic continuous curve, and every point of* H *is an end point both of* M *and of the curve* $(K+H)$.

THEOREM 18. *If* K *is any closed and connected subset of the set of all the cut points of a continuum* M, *then* K *contains at least one subcontinuum which belongs to the boundary of some single complementary domain of* M.

Proof. The complementary domains of M are countable. Let them be ordered $D_1, D_2, D_3, \cdots$, and let their respective boundaries be ordered $B_1, B_2, B_3, \cdots$. It is a consequence of a theorem of R. L. Moore's* that K is a subset of the point set $B_1+B_2+B_3+\cdots$. Let $A_1, A_2, A_3, \cdots$ denote the point sets common to $B_1, B_2, B_3, \cdots$ respectively, and to K. Then for every positive integer n, A_n is a closed point set. Now $K=A_1+A_2+A_3+\cdots$. It is well known that no continuum is expressible as the sum of a countable number of closed point sets each of which is totally disconnected. Hence for some positive integer i, A_i is not totally disconnected and therefore contains a continuum H. The continuum H belongs to B_i, the boundary of D_i.

THEOREM 19. *In order that the point* P *of a bounded continuum* M *should be a cut point of* M *it is necessary and sufficient that* P *should be a cut point of the boundary of some complementary domain of* M.

Proof. R. L. Moore has shown† that this condition is necessary. I shall show that it is sufficient. Suppose P is a cut point of the boundary N of a complementary domain D of a bounded continuum M.

Case I. Suppose D is bounded. Then let B denote the outer boundary‡ of D. R. L. Moore has shown§ that B has no cut point. Hence, $B-P$, in case P belongs to B, or B, in case P does not belong to B, must be a subset

---

* *Concerning the common boundary of two domains*, Fundamenta Mathematicae, vol. 6 (1924), pp. 203–213.

† Loc. cit.

‡ If D is a bounded domain, the *outer boundary* of D is the boundary of the unbounded complementary domain of $\overline{D}$. If D_1 and D_2 are mutually exclusive domains, the outer boundary of D_1 *with respect to* D_2 is the boundary of that complementary domain of $\overline{D_1}$ which contains D_2. Cf. R. L. Moore, *Concerning the separation of point sets by curves*, loc. cit., footnote to p. 475.

§ *Concerning the sum of a countable number of mutually exclusive continua in the plane*, loc. cit.

either of S_1 or of S_2, where S_1 and S_2 denote two mutually separated point sets into which, by hypothesis, N is divided by the omission of the point P. Suppose it belongs to S_1. Then let R denote the complementary domain of the continuum S_1+P which contains D. Since no point of S_2 belongs to S_1+P, and since every point of S_2 is a limit point of D, it follows that R contains S_2. Then S_2+P is a continuum which lies, except for the point P, wholly in R. By a theorem of R. L. Moore's,* there exists a simple closed curve J which contains P, encloses S_2, and lies, except for the point P, wholly in R. The curve J does not enclose or contain any point of $B-P$. Since J encloses S_2, it follows that $J-P$ contains a point of D. And since $J-P$ is connected and contains no point of N, then $J-P$ must be a subset of D. Hence, $J-P$ contains no point whatever of M. But S_2 belongs to the interior of J, and $B-P$ to the exterior of J, and J contains in common with M only the point P. It readily follows that P is a cut point of M.

Case II. Suppose D is unbounded. It is easily shown that there exists a ray r of an open curve which has exactly one point A, distinct from P, in common with N and lies except for the point A wholly in D. Now by hypothesis, $N-P$ is expressible as the sum of two mutually separated point sets S_1 and S_2, one of which, say S_1, contains the point A. The set $D-(r-A)$ is connected. Let R denote that complementary domain of the continuum S_1+r+P which contains $D-(r-A)$. The domain R is simply connected and contains S_2. Then by R. L. Moore's theorem quoted above there exists a simple closed curve J which encloses S_2, contains P, and lies except for the point P wholly in R. Just as in Case I it follows that $J-P$ is a subset of D and therefore contains no point of M. But J encloses S_2 and does not enclose or contain the point A. It follows that P is a cut point of M, and the theorem is proved.

THEOREM 20. *In order that the point P of a continuous curve M should be an end point of M it is sufficient (but not necessary) that P should be an end point of the boundary of some complementary domain of M.*

Proof. Let P denote a point of M which is an end point of N, the boundary of some complementary domain D of M. Suppose, contrary to this theorem, that P is not an end point of M. Then, by Theorem 12, M contains an arc APB having P as one of its interior points. Now either (1) each of the segments (AP) and (PB) of APB contains a point of N, or (2) one of these segments contains no point whatever of N. I will show that in either case P must belong to some simple closed curve of M. Suppose case (1) is true.

* *Concerning the separation of point sets by curves*, loc. cit., Theorem 3.

Then let X and Y denote points of N which belong to the segments AP and PB respectively of APB. Since P is not a cut point of N, it follows that $N-P$ contains an arc t from X to Y. The sum of the arcs t and APB contains a simple closed curve which contains P. Now suppose case (2) is true. Let S denote one of the segments (AP), (PB) of the arc APB which contains no point of N. Then S belongs to some complementary domain R of N. It follows from a theorem of R. L. Moore's* that the boundary of R is a simple closed curve which belongs to M. Clearly this curve must contain P. Hence, in any case, M contains a simple closed curve J which contains P. Let I and E denote the interior and exterior respectively of J. Then D is a subset either of I or of E, say of I. Let K denote the complementary domain of N which contains E. By R. L. Moore's theorem just cited, the boundary C of K is a simple closed curve which belongs to N. Clearly C must contain P. But by hypothesis P is an end point of N, and therefore, by Theorem 12, can belong to no simple closed curve of N. Thus the supposition that P is not an end point of M leads to a contradiction, and the theorem is proved.

THEOREM 21. *The set of all the end points of a continuous curve is totally disconnected.*†

Proof. Let K denote the set of all the end points of a continuous curve M. Suppose K contains a connected set H which consists of more than one point. Then from Theorem 12 and Theorem 7 it follows that every point of H must belong to the boundary of some complementary domain of M. Let D_1 denote a complementary domain of M which has the point A of H on its boundary. Now if H is a subset of the boundary of D_1, then by a theorem of R. L. Wilder's,‡ H is arcwise connected, and it easily follows that some point of H must be an interior point of some arc of M, contrary to Theorem 12. Hence, there exists a complementary domain D_2 of M which has on its boundary a point B of H which does not belong to the boundary of D_1. Let N denote the boundary of D_1. Let R denote the complementary domain of N which contains D_2. By R. L. Moore's theorem mentioned above, the boundary of R is a simple closed curve J. It is easily seen that J separates A from B. Hence, since H is connected, J must contain a point

---

* *Concerning continuous curves in the plane*, loc. cit., Theorem 4.

† My attention has been called to the fact that K. Menger has recently proved a proposition similar to Theorem 21. However, he uses the term *end point* in a different sense. Cf. K. Menger, *Grundzüge einer Theorie der Kurven*, Mathematische Annalen, vol. 95 (1925), pp. 272–306.

‡ Loc. cit., Theorem 20.

of H. But this is contrary to Theorem 12. It follows that K is totally disconnected.

THEOREM 22. *If K, H, and N respectively denote the set of all the cut points, end points, and simple closed curves of a continuous curve M, then $K+H+N=M$.*

Proof. Let P denote a point of M, if there be any, which is neither a cut point nor an end point of M. I shall show that P belongs to a simple closed curve of M and therefore belongs to N. Since P is not an end point of M, it follows by Theorem 12 that P is an interior point of some arc APB of M. And since P is not a cut point of M, it follows by R. L. Moore's theorem mentioned above that $M-P$ contains an arc t from A to B. On the arcs PA and PB of APB, in the order from P to A and from P to B respectively, let X and Y respectively denote the first points which belong to t. The simple closed curve formed by the arc XY of t plus the arc XPY of APB contains the point P. Hence, P belongs to N, and it follows that $K+H+N=M$.

THEOREM 23. *If N denotes the point set consisting of the sum of all the simple closed curves contained in a continuous curve M, then every connected subset of $M-N$ is arcwise connected.*

Proof. Let L denote any definite connected subset of $M-N$. It follows from Theorem 22 that every point of L is either a cut point or an end point of M. And since, by Theorem 21, the set of all the end points of M is totally disconnected, L must contain at least one point P which is a cut point of M. By the part of Theorem 19 established by R. L. Moore, P belongs to the boundary B of some complementary domain D of M. I will show that L is a subset of B. Suppose, on the contrary, that L contains a point Q which does not belong to B. Then Q lies in a complementary domain R of B. By R. L. Moore's theorem, the boundary J of R is a simple closed curve which belongs to B. Since L contains no point of N, J contains neither P nor Q. Now R is either the interior or the exterior of J. And if R is the interior [exterior] of J, then Q belongs to the interior [exterior] of J, and P belongs to the exterior [interior] of J. Hence, in any case P and Q are separated by J. Therefore, L contains a point of J, contrary to hypothesis. Thus the supposition that L contains a point which does not belong to B leads to a contradiction. Hence, L is a subset of B, and by a theorem of R. L. Wilder's* it follows that L is arcwise connected.

* Loc. cit.

THEOREM 24. *Under the same hypothesis as in Theorem* 23, *if* L *is any connected subset of* $M-N$, *then* $\overline{L}$ *is an acyclic continuous curve which belongs to the boundary of some single complementary domain of* M, *and every point of* $\overline{L}$ *is either a cut point or an end point of* M.

Proof. From the proof of Theorem 23 it follows that $\overline{L}$ belongs to the boundary B of some complementary domain D of M. Now since, by R. L. Wilder's theorem, every connected subset of B is arcwise connected, and since every point of $\overline{L}-L$ is a limit point of L by definition, it can easily be shown by methods identical with those used in the proof of Theorem 16 that every point of $\overline{L}-L$ is either a cut point or an end point of B. Now, by Theorem 19, every cut point of B is a cut point also of M; and by Theorem 20, every end point of B is an end point also of M. Hence, since by Theorem 22, every point of L is either a cut point or an end point of M, every point of $\overline{L}$ is either a cut point or an end point of M. By Theorem 9 and the above argument, it follows that $\overline{L}$ is an acyclic continuous curve which satisfies all the conditions of Theorem 24.

THEOREM 25. *If* M *is the complete boundary of two mutually exclusive domains* D_1 *and* D_2, *then no point of* M *is an end point of any continuum which contains* M.

Proof. It is sufficient to show that M contains no end point of itself. Suppose, on the contrary, that there exists a point P which is an end point of M. Then P belongs to no continuum of condensation of M. For let H be any subcontinuum of M which contains P. R. L. Moore has shown* that $M-H$ is connected. Hence, since, by supposition, P is an end point of M, P is not a limit point of $M-H$. Therefore P belongs to no continuum of condensation of M. By a theorem of R. L. Wilder's† it follows that P is accessible from each of the domains D_1 and D_2. Hence, if A and B are points of D_1 and D_2 respectively, there exist arcs AP and BP such that $AP-P$ and $BP-P$ are subsets of D_1 and D_2 respectively. Since for any continuum H of M which contains P, P is not a limit point of $M-H$, it can easily be shown that there exists an arc EFG from a point E of $AP-P$ to a point G of $BP-P$ which contains no point whatever of M. This is impossible, because E belongs to D_1, and G belongs to D_2, and D_1 and D_2 are mutually exclusive domains by hypothesis. Thus the supposition that M contains an end point leads to a contradiction, and the theorem is proved.

THEOREM 26. *No end point of a continuum* M *can be a boundary point of more than one complementary domain of* M.

* *Concerning the common boundary of two domains*, loc. cit., Theorem 2.

† Loc. cit., Theorem 2.

Proof. Suppose, on the contrary, that an end point P of M belongs to the boundary of each of two mutually exclusive domains D_1 and D_2 which are complementary to M. Let N denote the outer boundary of D_2 with respect to D_1. By a theorem of R. L. Moore's* N is the complete boundary of each of two mutually exclusive domains R_1 and R_2 which contain D_1 and D_2 respectively. And since P is a limit point both of R_1 and R_2, P must belong to N. But P is an end point of M, and by Theorem 25, it cannot belong to any point set which belongs to M and is the complete boundary of two mutually exclusive domains. Thus the supposition that P belongs to the boundary of more than one complementary domain of M leads to a contradiction.

THEOREM 27. *The collection G of all the continua (X) contained in the boundary M of a simply connected bounded domain D such that X is the complete boundary of some two mutually exclusive domains, is countable.*

Proof. Let K denote the unbounded complementary domain of M, and let B denote its boundary. For every element X of G, I shall define a domain R_x as follows. (1) When $X=B$, let $R_x=K$. (2) For every element (X) of G such that B is not a subset of X, the unbounded complementary domain of X contains D. For every such element X of G let R_x denote one *bounded* domain which has X as its boundary. (3) For every element (X) of G such that $B \neq X$ but such that B is a subset of X, it is true that X is the complete boundary of at least two *bounded* mutually exclusive domains, because for every such element X, the unbounded complementary domain of X is identical with K, and X is not the complete boundary of K. Not both of these bounded domains can contain points of D. Then for every such element X of G let R_x denote one of the bounded domains of which X is the boundary and which contains no point whatever of D.

Clearly, for every element X of G there corresponds a domain R_x as above defined. It is evident that for every element X, R_x is a complementary domain of M. It is well known that the collection T of all such domains (R_x) is countable. Since every element of G is the boundary of at least one domain of the collection T, it follows that G is countable.

I will remark here that Theorem 27 is a generalization of a theorem of R. L. Wilder's to the effect that the collection of all the simple closed curves contained in the boundary of a complementary domain of a continuous curve is countable.

---

* Loc. cit., Theorem 1.

THEOREM 28. *If K denotes the set of all the cut points of a bounded continuum M, G denotes the collection of all the continua (X) of M such that X is the complete boundary of two mutually exclusive domains, and T denotes the point set obtained by adding together all the point sets of the collection G, then the set of points common to K and T is countable.*

Proof. Let H denote the set of points common to K and T. Let the complementary domains of M be ordered $D_1, D_2, D_3, \cdots$, and their boundaries denoted by $B_1, B_2, B_3, \cdots$, respectively. Now by the part of Theorem 19 established by R. L. Moore, K is a subset of the point set $B_1+B_2+B_3+\cdots$. Hence, if for every i, A_i denotes the set of points common to H and B_i, then $H=A_1+A_2+A_3+\cdots$. I will show that for every positive integer i, A_i is a countable set of points. Let P denote a point of A_i. Then P belongs to some element X of G, and X is the complete boundary of two domains R_1 and R_2. One of these domains, say R_1, contains no point whatever of D_i. Let Y denote the outer boundary of D_i with respect to R_1. Then Y is an element of G which contains P and is a subset of B_i. Let G_i denote the collection of all those elements of G which are subsets of B_i. Then by Theorem 27, G_i is countable. It was just shown that every point of A_i belongs to some element of G_i. Since by R. L. Moore's theorem, no element of G contains any cut point of itself, it follows that no element of G_i contains more than a countable number of points of K. It follows, then, that A_i is countable, and therefore H is countable.

THEOREM 29. *If K denotes the set of all the cut points and N denotes the point set consisting of the sum of all the simple closed curves of a continuous curve, then the set of points common to K and N is countable.*

Theorem 29 is a corollary to Theorem 28.

THEOREM 30. *Every continuum M in a plane S is connected im kleinen at every one of its end points which is accessible from some point of $S-M$.*

Proof. Suppose P is any end point of M which is accessible from $S-M$. There exists an arc t having P as one of its extremities and such that $t-P$ is a subset of $S-M$. Suppose, contrary to this theorem, that M is not connected im kleinen at P. Then there exists a circle C' having center at P and such that every circle which is concentric with C' encloses a point X which belongs to M but which lies in no connected subset of M which contains P and is enclosed by C'. Let C be a circle concentric with C' and of diameter less than $\frac{1}{2}$ the diameter of C' and also less than $\frac{1}{4}$ the diameter of t. Then M contains a countable infinity of continua $M_\infty, M_1, M_2, M_3, \cdots$

such that (1) each of these continua has at least one point on C and is contained in C plus its interior, (2) no two of these continua have a point in common, and, indeed, no one of them, save possibly M_∞, is a proper subset of any connected point set common to M and to C plus its interior, (3) no point of the set $M_1+M_2+M_3+\cdots$ lies together with P in any connected subset of M which is enclosed by C', and (4) M_∞ contains the point P and is the sequential limiting set of the sequence of continua $M_1, M_2, M_3, \cdots$.* Let I denote M_∞ plus all the bounded complementary domains of M_∞. It is clear that I is a maximal connected subset of the closed point set $I+M_1+M_2+M_3+\cdots$, and that I neither separates the plane nor contains any point of $t-P$. Hence, by a theorem of R. L. Moore's† there exists a simple closed curve J which encloses I, contains no point of the point set $M_1+M_2+M_3+\cdots$, is a subset of the interior of C', and is such that its exterior contains at least one point A_0 of t. Let B (see Fig. 2) denote a point

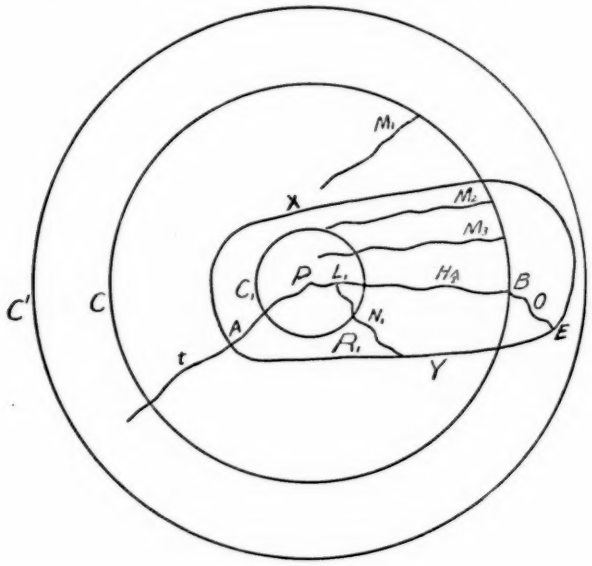

Fig. 2

* For indications of the proof of this statement see papers by R. L. Moore: *Continuous sets which have no continuous sets of condensation*, Bulletin of the American Mathematical Society, vol. 25 (1919), pp. 174–176, and *A characterization of Jordan regions by properties having no reference to their boundaries*, loc. cit.

† *Concerning the separation of point sets by curves*, loc. cit.

which is common to M_∞ and C. The point set M_∞ contains* a continuum H which is irreducible between P and B. Let H' denote the point set obtained by adding to H all of its bounded complementary domains. Now in the order from P to A_0 on t, let A denote the first point belonging to J. It is readily shown that there exists an arc BOE from B to a point E of J such that (BOE) is common to the interior of J and to the exterior of C. Let AXE and AYE respectively denote the two arcs of J from A to E. The continuum consisting of H' plus the arc PA of t plus the arc BOE divides the interior of J into just two domains D_1 and D_2. One of these domains, say D_1, has AXE in its boundary, and the other, D_2, has AYE in its boundary. It follows that one of these domains, say D_1, contains infinitely many of the continua $M_1, M_2, M_3, M_4, \cdots$.

Now let us consider the maximal connected subsets of $M-H$. It is evident that each of the continua $M_1, M_2, M_3, \cdots$ must belong to one such subset of $M-H$. And since P is an end point of M, it follows that no maximal connected subset of $M-H$ can contain more than a finite number of these continua. Hence, it is true that there exists an infinite sequence of distinct maximal connected subsets of $M-H$, each of which contains at least one of the continua $M_1, M_2, M_3, \cdots$. Let one such sequence be ordered $K_1, K_2, K_3, \cdots$. For every positive integer i, H contains at least one limit point of K_i. Let C_1 be a circle having P as center which lies entirely within J and is of diameter less than $\frac{1}{2}$ the diameter of C. From a theorem of Janiszewski's† it follows that H contains a continuum L_1 which contains P and a point of C_1 and which is the maximal connected subset of H which contains P and which belongs to C_1 plus its interior. By a theorem of Miss Mullikin's,‡ the continuum H contains a connected set Q which contains neither the point B nor any point of L_1 but which has B for a limit point and has at least one limit point in L_1. Now since H is irreducible between P and B, it readily follows that if H_1 denotes the point set $Q+B$, then $H=\overline{H}_1+L_1$. Since P is an end point of M, it follows (1) that P is not a limit point of H_1, and (2) that for not more than a finite number of positive integers (i) does $\overline{H}_1$ contain a limit point of K_i. Hence, there exists a positive integer n_1 such that $\overline{H}_1$ contains no limit point of K_{n_1}. Hence since H contains at least one limit point of K_{n_1}, L_1 must contain a point X_1 which is a limit point of K_{n_1}. Now from condition (3), above, which the sequence of sets $M_1, M_2, M_3, \cdots$ satisfies, it follows that for every positive integer i, K_i contains at least one point in common with J. Hence, by Miss Mullikin's

* Cf. Janiszewski, loc. cit.

† Loc. cit.

‡ These Transactions, vol. 24 (1922), pp. 144–162.

theorem mentioned above, $\overline{K}_{n_1}$ contains a connected set N_1^0 which contains no point of either of the continua L_1 and $J+BOE$ but is such that each of these continua contains at least one limit point of N_1^0. Clearly, N_1^0 is a subset either of D_1 or of D_2. And since D_1 contains infinitely many of the continua $M_1, M_2, M_3, \cdots$, of which only a finite number can contain points in common with N_1^0, it can be shown that N_1^0 cannot belong to D_1, and therefore, must belong to D_2. Let N_1 denote the point set obtained by adding to N_1^0 all of its limit points. It has already been shown that N_1 must be a subset of $D_2+L_1+$the arc $AYEOB$. It is evident that N_1 divides D_2 into at least two domains, one of which must have the arc AP of t in its boundary. Let R_1 denote the one which has AP in its boundary. It is clear, then, that no point of $\overline{H}_1-L_1\cdot\overline{H}_1$ is a limit point of R_1 and, therefore, that the boundary of R_1 is a subset of $N_1+L_1+$the arc $PAYEOB$.

Let C_2 be a circle concentric with C_1 which encloses and contains no point of either N_1 or H_1 and which is of diameter less than $\frac{1}{2}$ the diameter of C_1. Let L_2 be a subcontinuum of H which bears the same correspondence to C_2 as L_1 bears to C_1. Let sets H_2, K_{n_2}, N_2^0, and N_2 be selected and defined with respect to C_2 and L_2 just as the corresponding sets H_1, K_{n_1}, N_1^0, and N_1 were defined with respect to C_1 and L_1. Again, N_2^0 must be a subset of D_2. Hence, N_2 contains a point A_2 on the arc $AYEOB$. And since N_1 and N_2 can have no point in common, it can easily be shown that on $AYEOB$, in the order from A to B, A_2 precedes every point which belongs to N_1. Hence, N_2^0 is a subset of R_1. Let R_2 denote that complementary domain of the continuum $L_2+N_2+$the arc $PAYEOB$ which is a subset of R_1 and has the arc PA of t in its boundary. Again, $\overline{H}_2-L_2\cdot\overline{H}_2$ contains no limit point of R_2. This process may be continued indefinitely, and it follows that there exists an infinite sequence of continua $N_1, N_2, N_3, \cdots$, having the properties as above indicated. Also there exists a sequence of domains $R_1, R_2, R_3, \cdots$, such that for every positive integer n, R_n has the arc PA of t in its boundary, contains R_{n+1}, and contains $N_n+N_{n+1}+\cdots$. And there exist two sequences of connected point sets $L_1, L_2, L_3, \cdots$ and $H_1, H_2, H_3, \cdots$, such that for every positive integer n, $L_n+\overline{H}_n=H$, such that if r denotes the radius of C, then L_n contains P and is of diameter less than $3r/n$, and such that $\overline{H}_n-L_n\cdot\overline{H}_n$ contains no limit point whatever of R_n.

Let N denote the limiting set of the sequence of continua $N_1, N_2, N_3, \cdots$. It readily follows from the above properties of this sequence that N contains P but contains no other point whatever of H. The set N contains at least one point U of J. Now N is a continuum. Let N_u denote the maximal connected subset of $N-P$ which contains U. Then clearly P is a limit point of N_u. But P is an end point of M, and is, therefore, not a limit point

of any connected subset of $M-H$. Thus the supposition that M is not connected im kleinen at P leads to a contradiction, and the theorem is proved.

The following example demonstrates that the conclusion of Theorem 30 does not *necessarily* remain valid if the restriction that the end point of M in question shall be accessible from $S-M$ is removed. Let I be the straight line interval from (0,0) to (1,0). And for every integer n such that $n=(2)^i$, where i takes on all positive integral values from 1 to ∞, let L_n denote the broken line through the points $(1/n,0)$, $(1/n,-1/n)$, $(-1/n,-1/n)$, $(-1/n, 1/n)$, $(1,1/n)$, $(1,3/4n)$, and $(0,3/4n)$ in the order named. (See Fig. 3.)

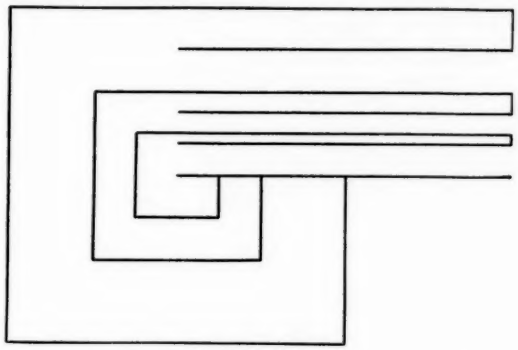

Fig. 3

If M denotes the continuum $I+L_1+L_2+L_3+\cdots$, and P denotes the point (0,0), then P is an end point of M, but M is not connected im kleinen at P.

THEOREM 31. *If a continuum M is irreducible between some pair of points A, B, then M is connected im kleinen at every one of its end points.*

Proof. Let P denote an end point of M. Let us first suppose that either $P\equiv A$ or $P\equiv B$, say $P\equiv B$. Then by Janiszewski's theorem mentioned above it follows that if C denotes any circle having P as center, then C encloses a subcontinuum H of M which consists of more than one point and which contains B but not A. From Miss Mullikin's theorem it follows immediately that $M-H$ contains a connected set N which contains A and which has at least one limit point in H. Since M is irreducible between A and P, clearly $M=H+\overline{N}$. And since P is an end point of M, P is not a limit point of $M-H$. Hence, there exists a circle K concentric with and within C which encloses no point of $M-H$. Any point of M which is interior to K lies together with P in a closed and connected subset of M which is enclosed by C, namely, in H itself. Hence, M is connected im kleinen at P.

Now in case neither $P\equiv A$ nor $P\equiv B$, then M is the sum of two continua K_a and K_b, irreducible between A and P, and B and P respectively. By the

above argument, both K_a and K_b are connected im kleinen at P. It follows that their sum, M, is connected im kleinen at P.

In his paper *Concerning the cut points of continuous curves and of other closed and connected point sets*, R. L. Moore proves the following results.

I. In order that a bounded continuum M should be an acyclic continuous curve it is necessary and sufficient that every subcontinuum of M should contain uncountably many points each of which is a cut point of M.

II. In order that the *continuous curve* M should contain no simple closed curve it is necessary and sufficient that if K denotes the set of all those points of M that are not cut points of M, then no subset of K disconnects M even in the weak sense.

In Theorem 32, below, I shall establish a generalization of R. L. Moore's result (II) quoted here.

THEOREM 32. *In order that the bounded continuum M should be an acyclic continuous curve it is necessary and sufficient that if K denotes the set of all those points of M which are not cut points of M, then no subset of K disconnects M even in the weak sense.*

Proof. The condition is sufficient. For suppose a bounded continuum M satisfies the condition but is not an acyclic continuous curve. Then by result (I), above, of R. L. Moore's, it follows that M contains a subcontinuum N which does not contain more than a countable number of cut points of M. Let A and B denote two points of N. By hypothesis, $M-[K-(A+B)]$ is connected in the strong sense. Hence, it contains a continuum H which contains A and B. Since every point of H, except possibly the points A and B, is a cut point of M, it follows by Theorem 9 that H is a continuous curve. Therefore, H contains an arc t from A to B. Since N contains not more than a countable number of cut points of M, there exist points E and F on t in the order A, E, F, B such that the interval EF of t contains no point whatever of N. Since N contains both A and B, it follows by Miss Mullikin's theorem that N contains a connected set Q containing no point of t and such that each of the intervals AE and FB of t contains at least one limit point of Q. But t is a continuum every point of which, save possibly two, is a cut point of M, and Q is a connected subset of $M-t$. Hence, by Theorem 11, t can contain at most one limit point of Q. Thus the supposition that M is not an acyclic continuous curve leads to a contradiction.

It follows by R. L. Moore's theorem II quoted above that the condition is necessary.

UNIVERSITY OF TEXAS,
AUSTIN, TEXAS

EXTREMALS AND TRANSVERSALITY OF THE GENERAL CALCULUS OF VARIATIONS PROBLEM OF THE FIRST ORDER IN SPACE*

BY

JESSE DOUGLAS†

I. INTRODUCTION

The ∞^2 extremals of the calculus of variations problem

$$\int F(x,y,y')dx = \text{minimum} \tag{1}$$

may be identified with an arbitrarily given system of ∞^2 curves in the plane,

$$y'' = \phi(x,y,y'), \tag{2}$$

by solving for F the partial differential equation

$$F_y - F_{y'x} - y'F_{y'y} - \phi F_{y'y'} = 0. \tag{3}$$

We may say, in short, that apart from consisting of ∞^2 curves the system of extremals of the general problem (1) has no properties.

It is different in three dimensions. To identify the ∞^4 extremals of

$$\int F(x,y,z,y',z')dx = \text{minimum} \tag{4}$$

with a given system of ∞^4 curves defined by the differential equations

$$\begin{aligned} y'' &= \phi(x,y,z,y',z'), \\ z'' &= \psi(x,y,z,y',z'), \end{aligned} \tag{5}$$

we must find F so as to satisfy simultaneously

$$\begin{aligned} F_y - F_{y'x} - y'F_{y'y} - z'F_{y'z} - \phi F_{y'y'} - \psi F_{y'z'} &= 0, \\ F_z - F_{z'x} - y'F_{z'y} - z'F_{z'z} - \phi F_{y'z'} - \psi F_{z'z'} &= 0. \end{aligned} \tag{6}$$

This can not be done, in general. Thus a system of extremals of a problem of the form (4) must have properties which distinguish it in the totality of curve families definable by equations of the form (5). To express these

* Presented to the Society, February 26, 1927; received by the editors in August, 1926.

† National Research Fellow in Mathematics.

properties analytically one would have to eliminate F from the equations (6) and their implications, a procedure requiring an appalling amount of calculation.* Nor could the result be given any perspicuous geometric interpretation.

Besides its system of extremals, every problem of the form (4) defines a second geometric object—its transversality. This is a certain correspondence between lineal elements (x, y, z, y', z') and surface elements (x, y, z, p, q) characterized by the following two properties: (A) a lineal element and its corresponding surface element have the same base point; (B) the ∞^2 extremals of (4) which are transverse to an arbitrary surface Σ admit ∞^1 transverse surfaces, between any two of which arcs over which $\int F(x, y, z, y', z')dx$ has the same value are intercepted on all these ∞^2 extremals.

Analytically, if the p, q of a surface element are defined as the coefficients in the equation of its plane when written in the form

$$\delta x + p\delta y + q\delta z = 0, \tag{7}$$

δx, δy, δz denoting relative coördinates as to x, y, z, the surface element transverse to the lineal element (x, y, z, y', z') is defined by

$$p = \frac{F_{y'}}{F - y'F_{y'} - z'F_{z'}}, \quad q = \frac{F_{z'}}{F - y'F_{y'} - z'F_{z'}}. \tag{8}$$

We shall use the term *transversality* to denote any correspondence between lineal and surface elements in which corresponding elements have the same base point. The general transversality is represented by two equations of the form

$$p = p(x,y,z,y',z'), \quad q = q(x,y,z,y',z'), \tag{9}$$

---

* Let D_r^5 denote the number of partial derivatives up to the rth order inclusive of a function of five independent variables, the function itself being included as its derivative of zeroth order. The formula for D_r^5 is $(r+1)(r+2)\cdots(r+5)\div 5!$. Then after forming all possible partial derivatives of the two equations (6) up to the rth order inclusive, we have $2D_r^5$ equations, which involve in a linear homogeneous way all the partial derivatives of F to the order $r+2$ inclusive except the following: F, $F_{y'}$, $F_{z'}$, the derivatives such as F_x, F_{xx}, $\cdots$, involving x alone up to the $(r+1)$th order, and the derivatives of order $r+2$, involving only x, y, z. If $d_{r+2}^3 (=(r+3)(r+4)\div 2)$ denote the number of partial derivatives of order $r+2$ of a function of three independent variables, the number of partial derivatives of F to be eliminated from the $2D_r^5$ equations is seen to be $D_{r+2}-d_{r+2}^3-(r+4)$. The number of equations remains less than the number of quantities to be eliminated until $r=11$, when these two numbers become respectively 8736, 8448. This indicates that we obtain by elimination 289 equations in the form of that number of determinants of order 8448 set equal to zero. These equations involve the derivatives of ϕ, ψ to the 11th order inclusive, of which there are 4368 for each function.

or inversely by

$$(9') \qquad y' = y'(x,y,z,p,q), \quad z' = z'(x,y,z,p,q),$$

where p, q and y', z' are arbitrary functions of their arguments.

Since (9) involves two arbitrary functions of five arguments, and (8) only one such function, it is obvious that a calculus of variations transversality must have special properties. These are expressed in analytic form by eliminating F between the two equations (8). A geometric interpretation of the resulting partial differential equation in p and q has been given by E. Kasner,* and is to the following effect.

In order that a transversality T belong to a problem in the calculus of variations, it is necessary and sufficient that it be definable as follows: With each point O of space associate, according to any (continuous, differentiable) law, a surface Σ_O. This may be done, and in the most general way, by laying off along each radius vector from O a segment OP whose length varies regularly from vector to vector, Σ_O being the locus of points P. Then define the transverse surface element σ to the lineal element λ of OP at O to be the element in O whose plane is parallel to the tangent plane of Σ_O at P.

A transversality T obtainable in the manner just described will be referred to in the statement of our main theorem as *of special character*.

We began by remarking that the characterization of the system of extremals of a variation problem (4) appears extremely difficult; and the transversality of such a problem has already been characterized. The object of the present paper is to characterize the composite of extremals and transversality.

In order to obtain a hint as to the nature of the characterization to be attempted, let us return to the property (B) of a variation transversality. The first part of (B), to the effect of the existence of ∞^1 transverse surfaces, is of purely geometric significance.

To each surface element σ of Σ let us construct the transverse lineal element λ; λ determines a definite extremal C, so that by varying σ over Σ we obtain the ∞^2 extremals transverse to Σ. Each of these has ∞^1 lineal elements, to each of which suppose constructed the transverse surface element. There arises in this way a field of ∞^3 surface elements (Pfaff equation). The significance of the first part of (B) is that this field of surface

* *Transversality for double integrals in the calculus of variations and for contact transformations*, an abstract of which appears in the Bulletin of the American Mathematical Society, vol. 21 (1914), pp. 71–72; see also p. 164.

elements is *integrable*; its ∞^3 elements can be resolved into ∞^1 surfaces (exact Pfaff equation).

It is evident that the same considerations continue to apply when Σ is any union—surface, curve, or point.

The situation just described leads to the formulation of the following theorem, whose proof is the goal of the present paper.

THEOREM. *If in connection with a family $\mathfrak{F}$ of ∞^4 curves in space definable analytically in the form* (5) *there exists a transversality T, necessarily of special character, such that the ∞^2 curves of $\mathfrak{F}$ which meet transversely an arbitrary base surface (union) Σ admit ∞^1 transverse surfaces, then $\mathfrak{F}$ must be the system of extremals of a calculus of variations problem whose transversality is T.*

This theorem is a generalization to the general problem (4) of a theorem of Kasner relative to problems of the important special form

$$\int \nu(x,y,z)ds = \int \nu(x,y,z)(1 + y'^2 + z'^2)^{1/2}dx = \text{minimum}. \tag{10}$$

It is readily shown that problems of the form (10) are characterized by the property that their transversality is orthogonality. Kasner proved* that if the ∞^2 curves of a given quadruply infinite family $\mathfrak{F}$ which meet an arbitrary surface Σ orthogonally admit ∞^1 orthogonal surfaces (form a normal congruence), then $\mathfrak{F}$ must be the system of extremals of a problem of the form (10).

Our theorem admits of extension to n dimensions without any difficulties besides those of calculation. In this form it includes a generalization by J. Lipka† of Kasner's theorem to any curved space of n dimensions. For in a curved as well as in a flat space a problem whose transversality is orthogonality is necessarily of the form $\int \nu ds = \text{minimum}$, where ν is a point function.

We may strengthen our theorem by requiring only that the property stated by it subsist for the following $5\infty^3$ base manifolds instead of for an arbitrary Σ: all points, all planes, any triple complex of curves. By the last we mean a system of $3\infty^3$ curves such that an arbitrary pencil of space contains three distinct directions which belong to curves of the system.

---

* *The theorem of Thomson and Tait and natural families of trajectories*, these Transactions, vol. 11 (1910), pp. 121–140. See also Princeton Colloquium Lectures, 1909, published 1913, §§35–44. The results of the cited paper are extended in a paper by the present author, *Normal congruences and quadruply infinite systems of curves in space*, these Transactions, vol. 26 (1924), pp. 68–100.

† *Some geometric investigations on the general problem of dynamics*, Proceedings of the American Academy of Arts and Sciences, vol. 55 (1920), pp. 285–322.

For instance, $3\infty^3$ curves in horizontal planes do not form a triple complex in this sense, since if any pencil with non-horizontal plane be chosen, the directions of curves of the given $3\infty^3$ which belong to this pencil coincide in its horizontal direction. Any three straight line complexes will serve.

II. Duality between the calculus of variations and infinitesimal contact transformations

We base our considerations on a certain duality between the calculus of variations and infinitesimal contact transformations (or one-parameter groups of contact transformations), which appears to have been first formulated by E. Vessiot.* The present section is devoted to a presentation of this duality in a form suitable for application later in this paper—a form entirely different from that adopted by Vessiot.

Let us consider first the general infinitesimal surface element transformation ϵ. It is defined by the equations

$$\begin{aligned} &\delta x = \delta t X(x,y,z,p,q), \quad \delta y = \delta t Y(x,y,z,p,q), \quad \delta t = \delta t Z(x,y,z,p,q), \\ &\delta p = \delta t P(x,y,z,p,q), \qquad\qquad \delta q = \delta t Q(x,y,z,p,q). \end{aligned} \tag{11}$$

Under ϵ the base point of each element σ, (x, y, z, p, q), receives an infinitesimal displacement in the direction $1:y':z'$, where

$$y' = \frac{Y(x,y,z,p,q)}{X(x,y,z,p,q)}, \quad z' = \frac{Z(x,y,z,p,q)}{X(x,y,z,p,q)}. \tag{12}$$

We shall call the lineal element (x, y, z, y', z') defined by (12) *transverse*† to σ. Thus every infinitesimal element transformation ϵ defines a certain transversality T_ϵ.

The transformation ϵ also defines a system of ∞^4 curves $\mathfrak{F}_\epsilon$ called its *path curves*. Allow ϵ to operate repeatedly on an element σ. Then the locus of the base point of σ is a curve termed a *path curve*‡ of ϵ. There are evidently ∞^4 such curves, since each of the ∞^5 elements σ of space may be used to generate one, while each path curve absorbs ∞^1 elements σ. The differential equations of the path curves $\mathfrak{F}_\epsilon$ are

---

* *Sur l'interprétation mécanique des transformations de contact infinitésimales*, Bulletin de la Société Mathématique de France, vol. 34 (1906), pp. 230–269.

† This term has been proposed in connection with contact (not general) transformations of lineal elements by Kasner, *The infinitesimal contact transformations of mechanics*, Bulletin of the American Mathematical Society, vol. 16 (1910), pp. 408–412.

‡ Vessiot uses the terms *conjugué, trajectoire* for *transverse, path curve*.

$$(13)\qquad \begin{aligned} y'' &= \left(\frac{Y}{X}\right)_x + \frac{Y}{X}\left(\frac{Y}{X}\right)_y + \frac{Z}{X}\left(\frac{Y}{X}\right)_z + \frac{P}{X}\left(\frac{Y}{X}\right)_p + \frac{Q}{X}\left(\frac{Y}{X}\right)_q, \\ z'' &= \left(\frac{Z}{X}\right)_x + \frac{Y}{X}\left(\frac{Z}{X}\right)_y + \frac{Z}{X}\left(\frac{Z}{X}\right)_z + \frac{P}{X}\left(\frac{Z}{X}\right)_p + \frac{Q}{X}\left(\frac{Z}{X}\right)_q, \end{aligned}$$

where in the second member we are to substitute for p, q the functions of x, y, z, y', z' which result by solving (12) for p, q. Formula (13) is obtained by expanding

$$y'' = \frac{\delta y'}{\delta x} = \frac{\delta\left(\frac{Y}{X}\right)}{\delta x}, \quad z'' = \frac{\delta z'}{\delta x} = \frac{\delta\left(\frac{Z}{X}\right)}{\delta x},$$

with the use of (11).

It is evident from (12) and (13) that the transversality and path curves of ϵ depend only on the ratios $X:Y:Z:P:Q$.

Conversely, if a transversality T and a quadruply infinite curve family $\mathfrak{F}$ are arbitrarily assigned, there exists an ϵ having T for its transversality and $\mathfrak{F}$ for its path curves, this ϵ being determined up to multiplication of its defining functions X, Y, Z, P, Q by a common factor $\rho(x, y, z, p, q)$, where ρ is an arbitrary function. For T determines Y/X, Z/X by (12); further, since $\mathfrak{F}$ is given, say in the form (5), the left members of (13) are equal to known functions ϕ, ψ of x, y, z, y', z', therefore by the substitution (12) to known functions of x, y, z, p, q; thus P/X, Q/X are uniquely determined, for the equations (13) are linear for P/X, Q/X, and their determinant, $(Y/X)_p(Z/X)_q - (Y/X)_q(Z/X)_p$ is not equal to zero, since, as we presume, the equations (12) which define T are solvable for p, q.

As is well known, an element transformation is called a *contact transformation* when and only when it converts every union of elements σ into a union of elements. The general infinitesimal contact transformation can be represented analytically in explicit form by means of its *characteristic function* $W(x, y, z, p, q)$ as follows:*

$$(14)\qquad \begin{aligned} &\delta x = \delta t(W - pW_p - qW_q), \quad && \delta y = \delta t W_p, \quad && \delta z = \delta t W_q, \\ &\delta p = \delta t(pW_z - W_y), && && \delta q = \delta t(qW_z - W_z). \end{aligned}$$

We shall refer to (14) as the infinitesimal contact transformation W. Its transversality is defined by

* This representation, based on the form (7) $\delta x + p\delta y + q\delta z = 0$ for the plane of an element σ, leads to more symmetrical formulas than the one given by Lie, *Theorie der Transformationsgruppen*, Leipzig, 1890, vol. 2, p. 252, and used by Vessiot, loc. cit., p. 230. Their representation is based on the form $p\delta x + q\delta y - \delta z = 0$ for the plane of σ.

$$y' = \frac{W_p}{W - pW_p - qW_q}, \qquad z' = \frac{W_q}{W - pW_p - qW_q}. \tag{15}$$

Now, a problem of first order in the calculus of variations is defined by a function $F(x, y, z, y', z')$ of the lineal elements λ of space; an infinitesimal contact transformation by a function $W(x, y, z, p, q)$ of the surface elements σ. To each F we can make correspond a W as follows. F being given, each element λ has a transverse element σ, defined by (11). Associate with σ the value W defined by the equation

$$WF = 1 + y'p + z'q. \tag{16}$$

In this way a function $W(x, y, z, p, q)$ is defined, which taken as characteristic function determines an infinitesimal contact transformation. We assert that the *transversality and path curves of W are identical respectively with the transversality and extremals of F.*

Conversely, let the infinitesimal contact transformation W be given. It causes to correspond to each element σ a transverse element λ, defined by (15). Associate with λ the value F defined by (16). *Then the variation problem defined by F has the same transversality as W, and its extremals are identical with the path curves of W.*

For the proof, we remark first that the equations (8) defining the transversality of F may be combined into

$$\frac{dF}{F} = \frac{pdy' + qdz'}{1 + y'p + z'q}, \tag{17}$$

where dF is formed on the supposition that the point (x, y, z) is fixed. Similarly, the equations (15) defining the transversality of W may be combined, on the same supposition, into

$$\frac{dW}{W} = \frac{y'dp + z'dq}{1 + y'p + z'q}. \tag{18}$$

By logarithmic differentiation of (16), (x, y, z) being regarded as fixed, we have

$$\frac{dF}{F} + \frac{dW}{W} = \frac{pdy' + qdz'}{1 + y'p + z'q} + \frac{y'dp + z'dq}{1 + y'p + z'q}. \tag{19}$$

It follows from (19) that (17) implies (18) and conversely. Thus if (y', z') and (p, q) correspond by the transversality of F, they correspond by the transversality of W, and conversely.

The extremals of F are defined by the differential equations

$$F_y - \frac{d}{dx}F_{y'} = 0, \tag{20}$$

$$F_z - \frac{d}{dx}F_{z'} = 0.$$

We have to prove that the same equations are satisfied by the path curves of W. From (8),

$$\frac{F_{y'}}{F} = \frac{p}{1 + y'p + z'q}, \quad \frac{F_{z'}}{F_{z'}} = \frac{q}{1 + y'p + z'q}; \tag{21}$$

therefore, by (16),

$$F_{y'} = \frac{p}{W}, \quad F_{z'} = \frac{q}{W}. \tag{22}$$

Thus, it has to be shown that along a path curve of W,

$$F_y - \frac{\delta}{\delta x}\left(\frac{p}{W}\right) = 0,$$

$$F_z - \frac{\delta}{\delta x}\left(\frac{q}{W}\right) = 0. \tag{23}$$

We have

$$\delta\left(\frac{p}{W}\right) = \frac{W\delta p - p\delta W}{W^2}$$

$$= \frac{1}{W^2}\{W\delta p - p(W_x\delta x + W_y\delta y + W_z\delta z + W_p\delta p + W_q\delta q)\}. \tag{24}$$

By means of (14) the bracket reduces to $-WW_p\delta t$, so that

$$\delta\left(\frac{p}{W}\right) = -\frac{W_y}{W}\delta t. \tag{25}$$

Again, by (14),

$$\delta x = \delta t(W - pW_p - qW_q). \tag{26}$$

Making the substitutions (15) in (16), we find

$$F = \frac{1}{W - pW_p - qW_q}; \tag{27}$$

hence

$$\delta x = \frac{1}{F}\,\delta t. \tag{28}$$

Substituting (25), (28) in the first equation (23), we obtain

$$\frac{F_y}{F} + \frac{W_y}{W} = 0. \tag{29}$$

Similarly the second equation (23) can be made to depend upon

$$\frac{F_z}{F} + \frac{W_z}{W} = 0. \tag{30}$$

The identity of the extremals of F and the path curves of W is established if we can prove (29), (30). To this end, let us form the logarithmic derivative of (16) partially with respect to y, from the point of view that the p, q in W and in the right member are functions of x, y, z, y', z' by means of the common transversality of W and F. We find

$$\frac{F_y}{F} + \frac{W_y}{W} + \frac{W_p}{W}\frac{\partial p}{\partial y} + \frac{W_q}{W}\frac{\partial q}{\partial y} = \frac{y'}{1 + y'p + z'q}\frac{\partial p}{\partial y} + \frac{z'}{1 + y'p + z'q}\frac{\partial q}{\partial y}. \tag{31}$$

By (15) the terms in $\partial p/\partial y$, $\partial q/\partial y$ cancel, giving

$$\frac{F_y}{F} + \frac{W_y}{W} = 0,$$

or (29). Similarly (30) can be proved. Thus the duality between F and W is established.

Suppose now that an infinitesimal element transformation ϵ is applied repeatedly to the ∞^2 elements of an arbitrary surface Σ. A field of ∞^3 surface elements is generated which may or may not be integrable. With the developments of the present section in mind we may reformulate our main theorem stated in the preceding section, as follows:

THEOREM. *If the field of ∞^3 surface elements generated by applying a given infinitesimal element tranformation ϵ with infinite repetition to the elements of an arbitrary surface (union) Σ is integrable, then ϵ must be proportional to an infinitesimal contact transformation W.*

By the last phrase we mean that the defining functions X, Y, Z, P, Q of ϵ are proportional to those of an infinitesimal contact transformation as they appear in (14).

The sequel is devoted to a proof of the theorem in the form just stated.

III. Conditions that an infinitesimal element transformation be proportional to an infinitesimal contact transformation

We prove in this section that, given X, Y, Z, P, Q, necessary and sufficient conditions that there exist a ρ and a W such that

$$(32)\qquad \begin{aligned} X &= \rho(W - pW_p - qW_q), & Y &= \rho W_p, & Z &= \rho W_q, \\ P &= \rho(pW_x - W_y), & Q &= \rho(qW_x - W_z), \end{aligned}$$

are the following, where

$$(33)\qquad M = X + pY + qZ:^*$$

$$(34)\quad \text{(a)}\qquad \left(\frac{Y}{M}\right)_q = \left(\frac{Z}{M}\right)_p,$$

$$\text{(b)}\qquad \left(\frac{P}{M}\right)_q = p\left(\frac{Z}{M}\right)_x - \left(\frac{Z}{M}\right)_y,$$

$$\text{(c)}\qquad \left(\frac{Q}{M}\right)_p = q\left(\frac{Y}{M}\right)_x - \left(\frac{Y}{M}\right)_z,$$

$$\text{(d)}\qquad p\left(\frac{Y}{M}\right)_x - \left(\frac{Y}{M}\right)_y - \left(\frac{P}{M}\right)_p = q\left(\frac{Z}{M}\right)_x - \left(\frac{Z}{M}\right)_z - \left(\frac{Q}{M}\right)_q,$$

$$\text{(e)}\qquad q\left(\frac{P}{M}\right)_x - \left(\frac{P}{M}\right)_z = p\left(\frac{Q}{M}\right)_x - \left(\frac{Q}{M}\right)_y.$$

To prove the conditions (34) necessary, or that (32) implies (34), we first combine (32) and (33) to obtain

$$(35)\qquad M = \rho W.$$

Dividing (32) by (35), we have, with the notation

$$(36)\qquad V = \log W,$$

$$(37)\qquad \begin{aligned} \frac{X}{M} &= 1 - pV_p - qV_q, & \frac{Y}{M} &= V_p, & \frac{Z}{M} &= V_q, \\ \frac{P}{M} &= pV_x - V_y, & \frac{Q}{M} &= qV_x - V_z. \end{aligned}$$

The equations (34) were derived by elimination of V from (37), and their necessity can be verified by direct substitution.

* We assume $M \neq 0$, that is, that the lineal element transverse to any given surface element does not lie in the plane of that element.

To establish the sufficiency of the conditions (34), we begin by observing that (34) (a) implies the existence of a $\sigma(x, y, z, p, q)$ such that

$$\frac{Y}{M} = \sigma_p, \qquad \frac{Z}{M} = \sigma_q, \tag{38}$$

and (34) (e) the existence of a $\tau(x, y, z, p, q)$ such that

$$\frac{P}{M} = p\tau_x - \tau_y, \qquad \frac{Q}{M} = q\tau_x - \tau_z, \tag{39}$$

for (34) (e) can be written

$$\left(q\frac{\partial}{\partial x} - \frac{\partial}{\partial z}\right)\frac{P}{M} = \left(p\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right)\frac{Q}{M}, \tag{40}$$

and the linear partial differential operators involved have coefficients constant as to the variables of differentiation.*

Substituting the values (38), (39) in (34) (b), (c), (d), we obtain

$$p\tau_{qx} - \tau_{qy} = p\sigma_{qx} - \sigma_{qy}, \tag{41}$$

$$q\tau_{px} - \tau_{pz} = q\sigma_{px} - \sigma_{pz}, \tag{42}$$

$$(p\sigma_{pz} - \sigma_{py}) - (p\tau_{px} - \tau_{py}) = (q\sigma_{qx} - \sigma_{qz}) - (q\tau_{qx} - \tau_{qz}). \tag{43}$$

These can be written as equations in

$$\omega = \tau - \sigma \tag{44}$$

as follows:

$$\left(p\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right)\omega_q = 0, \tag{45 (a)}$$

$$\left(q\frac{\partial}{\partial x} - \frac{\partial}{\partial z}\right)\omega_p = 0, \tag{(b)}$$

$$\left(p\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right)\omega_p = \left(q\frac{\partial}{\partial x} - \frac{\partial}{\partial z}\right)\omega_q, \tag{(c)}$$

---

* The general solution of

$$\sum b_i \frac{\partial}{\partial x_i}\phi = \sum a_i \frac{\partial}{\partial x_i}\psi,$$

where the a_i, b_i are constants, is

$$\phi = \sum a_i \frac{\partial}{\partial x_i}\omega, \quad \psi = \sum b_i \frac{\partial}{\partial x_i}\omega,$$

where ω is an arbitrary function of the variables x.

of which we have to find the general solution. Equations (45) (a), (b) imply respectively

$$\omega_q = \psi(\xi, z, p, q), \tag{46}$$

$$\omega_p = \phi(\xi, y, p, q), \tag{47}$$

where

$$\xi = x + py + qz. \tag{48}$$

Substituting in (45) (c), we obtain

$$p\phi_\xi - p\phi_\xi - \phi_y = q\psi_\xi - q\psi_\xi - \psi_z,$$

or

$$\phi_y = \psi_z, \tag{49}$$

whose common value, regarding ξ, y, z, p, q for the moment as the independent variables of our problem, must be an $f(\xi, p, q)$, for ϕ_y is free of z and ψ_z free of y. It follows that

$$\phi = \omega_p = f(\xi, p, q)y + g(\xi, p, q), \tag{50}$$

$$\psi = \omega_q = f(\xi, p, q)z + h(\xi, p, q). \tag{51}$$

We apply now the condition $\omega_{pq} = \omega_{qp}$, which gives

$$f_\xi yz + f_q y + g_\xi z + g_q = f_\xi yz + f_p z + h_\xi y + h_p\,; \tag{52}$$

and therefore, since (52) must hold identically in ξ, y, z, p, q,

$$f_q = h_\xi, \qquad g_\xi = f_p, \qquad g_q = h_p. \tag{53}$$

This implies the existence of a $\lambda(\xi, p, q)$ such that

$$f = \lambda_\xi, \qquad g = \lambda_p, \qquad h = \lambda_q. \tag{54}$$

Thus (50), (51) become

$$\omega_p = \lambda_\xi y + \lambda_p, \tag{55}$$

$$\omega_q = \lambda_\xi z + \lambda_q. \tag{56}$$

Going back to x, y, z, p, q as independent variables in λ, the second members of (55), (56) are respectively $\partial\lambda/\partial p$, $\partial\lambda/\partial q$; hence the general solution of these equations is

$$\omega = \lambda(\xi, p, q) - \mu(x, y, z), \tag{57}$$

where μ is an arbitrary function of its arguments.

By this and (44) we have

$$\tau - \lambda(\xi, p, q) = \sigma - \mu(x, y, z). \tag{58}$$

Since, as is readily seen, the modification of τ by the subtractive $\lambda(\xi, p, q)$ does not affect the values of the second members of (39), nor the modification of σ by the subtractive $\mu(x, y, z)$ the values of the second members of (38), if we denote by $V(x, y, z, p, q)$ the common value of the two members of (58), we have

$$\begin{aligned} \frac{Y}{M} &= V_p, & \frac{Z}{M} &= V_q, \\ \frac{P}{M} &= pV_z - V_y, & \frac{Q}{M} &= qV_z - V_x, \end{aligned} \tag{59}$$

to which we may adjoin

$$\frac{X}{M} = 1 - pV_p - qV_q \tag{59'}$$

by the definition (33) of M.

The existence of this V established, we can construct

$$W = e^V, \qquad \rho = Me^{-V}. \tag{60}$$

Substituting

$$V = \log W, \qquad M = \rho W, \tag{61}$$

in (59), (59'), we obtain the system (32). Thus the proof is completed that the conditions (34) are necessary and sufficient for the existence of a W and a ρ verifying (32).

IV. Analytic solution of the problem

Our problem is now the following: To prove that if a given infinitesimal element transformation ϵ

$$\begin{aligned} \frac{\delta x}{\delta t} &= X(x,y,z,p,q), & \frac{\delta y}{\delta t} &= Y(x,y,z,p,q), & \frac{\partial z}{\delta t} &= Z(x,y,z,p,q), \\ \frac{\delta p}{\delta t} &= P(x,y,z,p,q), & \frac{\delta q}{\delta t} &= Q(x,y,z,p,q), \end{aligned} \tag{62}$$

applied repeatedly to the elements of an *arbitrary* surface Σ generates a field of ∞^3 elements which is integrable, then ϵ must obey the system (34).

Denoting by the subscript 0 quantities relative to Σ, we write the equation of that surface in the form

$$x_0 = -\phi(y_0, z_0). \tag{63}$$

Its ∞^2 surface elements are expressed in terms of the two parameters y_0, z_0 by the equations

$$x_0 = -\phi(y_0,z_0), \quad y_0 = y_0, \quad z_0 = z_0, \quad p_0 = \phi_{y_0}(y_0,z_0), \quad q_0 = \phi_{z_0}(y_0,z_0). \tag{64}$$

We may regard (62) as a system of differential equations for x, y, z, p, q as functions of t. Let the uniquely determined solution of this system which takes the values (64) for $t=0$ be denoted by

$$\begin{aligned} &x = x(t,y_0,z_0), \quad y = y(t,y_0,z_0), \quad z = z(t,y_0,z_0), \\ &p = p(t,y_0,z_0), \quad q = q(t,y_0,z_0). \end{aligned} \tag{65}$$

These equations represent in terms of the three parameters t, y_0, z_0 the field of ∞^3 elements generated from Σ by means of ϵ. The condition that this field be integrable is

$$(pxy) - (qzx) + p(qyz) - q(pyz) = 0,^* \tag{66}$$

where the parentheses denote jacobians as to t, y_0, z_0.

Using (62), this can be expanded into

$$\begin{aligned} P(xy) + X(yp) + Y(px) - Q(zx) - Z(xq) - X(qz) \\ + pQ(yz) + pY(zq) + pZ(qy) - qP(yz) - qY(zp) - qZ(py) = 0, \end{aligned} \tag{67}$$

where the parentheses denote determinants in the matrix

$$\begin{vmatrix} \frac{\partial x}{\partial y_0} & \frac{\partial y}{\partial y_0} & \frac{\partial z}{\partial y_0} & \frac{\partial p}{\partial y_0} & \frac{\partial q}{\partial y_0} \\ \frac{\partial x}{\partial z_0} & \frac{\partial y}{\partial z_0} & \frac{\partial z}{\partial z_0} & \frac{\partial p}{\partial z_0} & \frac{\partial q}{\partial z_0} \end{vmatrix}. \tag{68}$$

We introduce to represent these determinants ten symbols ω as follows:

$$\begin{aligned} &\omega_1 = (xy), \quad \omega_2 = (xz), \quad \omega_3 = (xp), \quad \omega_4 = (xq), \quad \omega_5 = (yz), \\ &\omega_6 = (yp), \quad \omega_7 = (yq), \quad \omega_8 = (zp), \quad \omega_9 = (zq), \quad \omega_{10} = (pq). \end{aligned} \tag{69}$$

Formula (67) can then be abbreviated to

$$\sum A_i \omega_i = 0 \tag{70}$$

by means of the further notation

$$\begin{aligned} &A_1 = P, \quad A_2 = Q, \quad A_3 = -Y, \quad A_4 = -Z, \quad A_5 = pQ - qP, \\ &A_6 = X + qZ, \quad A_7 = -pZ, \quad A_8 = -qY, \quad A_9 = X + pY, \quad A_{10} = 0. \end{aligned} \tag{71}$$

* This condition is obtained by making the change of variables (65) in the Pfaff equation $dx + p\,dy + q\,dz = 0$, and then applying the usual criterion for exactness to the resulting Pfaffian in t, y_0, z_0.

In (70) and everywhere hereafter the index runs from 1 to 10.

Denoting by means of a bar the partial derivatives of x, y, z, p, q as to y_0, we see that these obey the equations of variation of (62):

$$\begin{aligned} \bar{x}' &= X_x\bar{x} + X_y\bar{y} + X_z\bar{z} + X_p\bar{p} + X_q\bar{q}, \\ &\cdots\cdots\cdots\cdots \\ \bar{q}' &= Q_x\bar{x} + Q_y\bar{y} + Q_z\bar{z} + Q_p\bar{p} + Q_q\bar{q}, \end{aligned} \tag{72}$$

obtained by substituting (65) in (62) and differentiating the resulting identities in t, y_0, z_0 partially as to y_0. The accent denotes differentiation as to t with y_0, z_0 held fixed, that is, differentiation along a path curve of ϵ.

Evidently the partial derivatives of x, y, z, p, q as to z_0 obey the same equations of variation.

From this it follows that the determinants ω obey a system of linear differential equations. Consider for instance $\omega_1 = (xy)$. We have

$$\begin{aligned} \omega_1' &= (x'y) + (xy') \\ &= X_x(xy) + X_z(zy) + X_p(py) + X_q(qy) \\ &\quad + Y_y(xy) + Y_z(xz) + Y_p(xp) + Y_q(xq) \end{aligned} \tag{73}$$

by (72), or

$$\omega_1' = (X_x + Y_y)\omega_1 + Y_z\omega_2 + Y_p\omega_3 + Y_q\omega_4 - X_z\omega_5 - X_p\omega_6 - X_q\omega_7. \tag{74}$$

Similarly each ω' can be expressed as a homogeneous linear form in the ω's. Carrying the calculations through, we obtain the system of linear differential equations represented by the following scheme (75):

	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9	ω_{10}
ω_1'	X_x+Y_y	Y_z	Y_p	Y_q	$-X_z$	$-X_p$	$-X_q$			
ω_2'	Z_y	X_x+Z_z	Z_p	Z_q	X_y			$-X_p$	$-X_q$	
ω_3'	P_y	P_z	X_x+P_p	P_q		X_y		X_z		$-X_q$
ω_4'	Q_y	Q_z	Q_p	X_x+Q_q			X_y		X_z	X_p
ω_5'	$-Z_x$	Y_x			Y_y+Z_z	Z_p	Z_q	$-Y_p$	$-Y_q$	
ω_6'	$-P_x$		Y_x		P_z	Y_y+P_p	P_q	Y_z		$-Y_q$
ω_7'	$-Q_x$			Y_x	Q_z	Q_p	Y_y+Q_q		Y_z	Y_p
ω_8'		$-P_x$	Z_x		$-P_y$	Z_y		Z_z+P_p	P_q	$-Z_q$
ω_9'		$-Q_x$		Z_x	$-Q_y$		Z_y	Q_p	Z_z+Q_q	Z_p
ω_{10}'			$-Q_x$	P_x		$-Q_y$	P_y	$-Q_z$	P_z	P_p+Q_q

As a solution of the system (75), the ω's will be completely determined if we know their values for $t=0$, that is, on the base surface Σ. The values on Σ of the elements of the matrix (68) are evidently the following:

$$\begin{vmatrix} -p_0 & 1 & 0 & r_0 & s_0 \\ -q_0 & 0 & 1 & s_0 & t_0 \end{vmatrix} \tag{76}$$

where r_0, s_0, t_0 denote the second partial derivatives of the function ϕ which defines Σ.

Hence the values of the ω's on Σ, which values we denote by the symbols α, are

$$\begin{aligned} &\alpha_1 = q_0, \quad \alpha_2 = -p_0, \quad \alpha_3 = -p_0s_0 + q_0r_0, \quad \alpha_4 = -p_0t_0 + q_0s_0, \\ &\alpha_5 = 1, \quad \alpha_6 = s_0, \quad \alpha_7 = t_0, \quad \alpha_8 = -r_0, \\ &\alpha_9 = -s_0, \quad \alpha_{10} = r_0t_0 - s_0^2. \end{aligned} \tag{77}$$

The necessary and sufficient condition that the field of elements generated from Σ by means of ϵ be integrable is that the ten ω's which satisfy (75) *and take for $t=0$ the values* (77) *shall obey the linear relation* (70), *and that identically in the parameters y_0, z_0.*

The analytic expression of this condition is a problem which we have already solved in a previous paper.*

Let us denote by Ω the left member of (70):

$$\Omega = \sum_i A_i\omega_i.$$

The ω's are functions of t, y_0, z_0, and the A's functions of x, y, z, p, q, which after the substitution (65) also become functions of t, y_0, z_0. Assuming, as we do, that all functions involved in our considerations are analytic, the necessary and sufficient condition for the identical vanishing of Ω is that Ω together with its derivatives of all orders vanish for $t=0$, identically in y_0, z_0.

Using after each successive differentiation of Ω the system (75), which we abbreviate to

$$\omega_i' = \sum_j p_{ij}\omega_j, \tag{78}$$

we find that each derivative Ω', Ω'', $\cdots$ is, like Ω itself, a homogeneous linear form in the ω's. We obtain, in fact,

$$\Omega = \sum_i A_i\omega_i, \quad \Omega' = \sum_i B_i\omega_i, \quad \Omega'' = \sum_i C_i\omega_i, \ \cdots, \tag{79}$$

* *Normal congruences* . . . , previously cited; see §5 of that paper.

where each set of ten symbols $B, C, \cdots$ is derived from its predecessor the recurrence formulas

$$
\begin{aligned}
B_i &= A_i' + \sum_j p_{ji} A_j, \\
C_i &= B_i' + \sum_j p_{ji} B_j, \\
&\cdots\cdots
\end{aligned}
\tag{80}
$$

in which the accent denotes d/dt along a path curve of ϵ, that is

$$
' \equiv X\frac{\partial}{\partial x} + Y\frac{\partial}{\partial y} + Z\frac{\partial}{\partial z} + P\frac{\partial}{\partial p} + Q\frac{\partial}{\partial q}.
$$

The quantities $B, C, \cdots$ are thus expressions of higher and higher order in the partial derivatives of X, Y, Z, P, Q.

In demanding that $\Omega, \Omega', \Omega'', \cdots$ all vanish for $t=0$, we get the infinite sequence of conditions

$$
\sum_i A_i\alpha_i = 0, \qquad \sum_i B_i\alpha_i = 0, \qquad \sum_i C_i\alpha_i = 0, \cdots,
\tag{81}
$$

where the $A, B, C, \cdots$ are the values for $t=0$ of the $A, B, C, \cdots$ of (79); in other words, the $A, B, C, \cdots$ of (81) are the values on Σ of the same letters of (79), and are therefore to be derived from the latter by attaching the subscript 0 to x, y, z, p, q.

Substituting the values (77) of the α's in (81), we obtain, after dropping the subscript 0, the following infinite system of partial differential equations of the Monge-Ampère type:

$$
\mathfrak{B}_0(rt - s^2) + \mathfrak{B}_1 r + \mathfrak{B}_2 s + \mathfrak{B}_3 t + \mathfrak{B}_4 = 0,
\tag{82$_1$}
$$

$$
\mathfrak{C}_0(rt - s^2) + \mathfrak{C}_1 r + \mathfrak{C}_2 s + \mathfrak{C}_3 t + \mathfrak{C}_4 = 0,
\tag{82$_2$}
$$

$$\cdots\cdots$$

where the symbols $\mathfrak{B}, \mathfrak{C}, \cdots$ are derived respectively from $B, C, \cdots$ by the same formulas, typified by

$$
\begin{aligned}
\mathfrak{B}_0 &= B_{10}, \\
\mathfrak{B}_1 &= qB_3 - B_8, \\
\mathfrak{B}_2 &= -pB_3 + qB_4 + B_6 - B_8, \\
\mathfrak{B}_3 &= -pB_4 + B_7, \\
\mathfrak{B}_4 &= qB_1 - pB_2 + B_5.
\end{aligned}
\tag{83}
$$

The equation (82_0), $\mathfrak{A}_0(rt-s^2)+\cdots$, disappears identically, that is, each $\mathfrak{A}=0$, as a consequence of the fact that Σ itself is always an integral surface of the field of elements which it generates.

A surface Σ gives rise to an integrable element-field when and only when it obeys the infinite system (82).

If an *arbitrary* surface Σ gives an integrable element-field, each equation (82) must disappear identically. In particular (82_1) must do so, that is,

$$\mathfrak{B}_0 = 0, \qquad \mathfrak{B}_1 = 0, \qquad \mathfrak{B}_2 = 0, \qquad \mathfrak{B}_3 = 0, \qquad \mathfrak{B}_4 = 0. \tag{84}$$

Let us calculate the explicit values of the B's by carrying out the operations indicated in (80). We find, with reference to (71) and the scheme (75),

$$\begin{aligned}
B_1 &= \{PX_x + (qP - pQ)Z_x - qZP_x + pZQ_x\} + (PY_y + QZ_y - ZQ_y) \\
&\quad + ZP_x + PP_p + QP_q, \\
B_2 &= \{QX_x + (pQ - qP)Y_x + qYP_x - pYQ_x\} + YQ_y \\
&\quad + (PY_x + QZ_x - YP_x) + PQ_p + QQ_q, \\
B_3 &= (-YX_x + qZY_x - qYZ_x) - YY_y - ZY_x \\
&\quad + (QZ_p - YP_p - ZQ_p) - QY_q, \\
B_4 &= (-ZX_x - pZY_x + pYZ_x) - YZ_y - ZZ_x \\
&\quad - PZ_p + (PY_q - YP_q - ZQ_q), \\
B_5 &= (-qXP_x + pXQ_x) + \{QX_y + (pQ - qP)Y_y - XQ_y\} \\
&\quad + \{-PX_x + (pQ - qP)Z_x + XP_x\} \\
&\quad + (-qPP_p + pPQ_p) + (-qQP_q + pQQ_q), \\
B_6 &= ZQ + (XX_x + qXZ_x) + (X + qZ)Y_y + (ZX_x + qZZ_x) \\
&\quad + \{pQZ_p + (X + qZ)P_p - pZQ_p\} + (QX_q + qQZ_q), \\
B_7 &= -ZP - pXZ_x + (-ZX_y - pZY_y + XZ_y) - pZZ_x \\
&\quad - pPZ_p + \{-PX_q - qPZ_q + (X + qZ)P_q - pZQ_q\}, \\
B_8 &= -YQ - qXY_x - qYY_y + (-YX_x + XY_x - qYZ_x) \\
&\quad + \{-QX_p - pQY_p - qYP_p + (X + pY)Q_p\} - qQY_q,
\end{aligned} \tag{85}$$

$$B_9 = YP + (XX_x + pXY_x) + (YX_y + pYY_y) + (X + pY)Z_z + (PX_p + pPY_p) + \{qPY_q - qYP_q + (X + pY)Q_q\},$$

$$B_{10} = \{ -ZX_p - pZY_p + (X + pY)Z_p\} + \{YX_q - (X + qZ)Y_q + qYZ_q\}$$

From these we can then determine by means of (83) the values of the symbols $\mathfrak{B}$; they can be written, using $M = X + pY + qZ$,

$$\begin{aligned}
\mathfrak{B}_0 &= (MZ_p - ZM_p) - (MY_q - YM_q),\\
\mathfrak{B}_1 &= q(MY_x - YM_x) - (MY_z - YM_z) - (MQ_p - QM_p),\\
\mathfrak{B}_2 &= -p(MY_x - YM_x) + q(MZ_x - ZM_x) + (MY_y - YM_y)\\
&\quad - (MZ_z - ZM_z) + (MP_p - PM_p) - (MQ_q - QM_q),\\
\mathfrak{B}_3 &= -p(MZ_x - ZM_x) + (MZ_y - ZM_y) + (MP_q - PM_q),\\
\mathfrak{B}_4 &= -q(MP_x - PM_x) + p(MQ_x - QM_x)\\
&\quad - (MQ_y - QM_y) + (MP_z - PM_z).
\end{aligned} \tag{86}$$

After division by M^2, the five conditions $\mathfrak{B} = 0$ are seen to be identical with the system (34), *previously proved necessary and sufficient in order that ϵ be proportional to an infinitesimal contact transformation. Hence our main theorem.*

V. Stronger forms of statement

The theorem was derived from the conditions $\mathfrak{B} = 0$ alone, instead of from the complete infinite system $\mathfrak{B} = 0$, $\mathfrak{C} = 0$, $\cdots$. The identical disappearance of only the first equation (82) signifies that the element-field generated from Σ by ϵ has, besides Σ, an integral surface consecutive to Σ. We may therefore strengthen our theorem by substituting this requirement in its hypothesis instead of that of the existence of ∞^1 integral surfaces.

The other stronger form stated in the introduction, in which certain $5\infty^3$ manifolds are substituted for the ∞^∞ manifolds Σ, results from the following observations.

We may regard points and planes as spheres of radius zero and infinity respectively. For a sphere of radius R we have

$$\frac{r}{1 + p^2} = \frac{s}{pq} = \frac{t}{1 + q^2} = \frac{(1 + p^2 + q^2)^{1/2}}{R}. \tag{87}$$

Substituting the resulting values of r, s, t in (82_1), we obtain a quadratic equation in R. If this equation is satisfied by $R=0$ and $R=\infty$, the constant term and the coefficient of R^2 must reduce to zero, that is

$$\mathfrak{B}_0 = 0, \qquad \mathfrak{B}_4 = 0. \tag{88}$$

Again, for a curve regarded as the limit as $R\to 0$ of a canal surface generated by a sphere of radius R, we can calculate that r, s, t become infinite in the ratios

$$r:s:t = z'^2:-y'z':y'^2 \tag{89}$$

where $1:y':z'$ defines the direction of the curve at the particular point under consideration.

Under (88) the equation (82_1) reduces to

$$\mathfrak{B}_1 r + \mathfrak{B}_2 s + \mathfrak{B}_3 t = 0. \tag{90}$$

This is satisfied by a curve C when and only when

$$\mathfrak{B}_1 z'^2 - \mathfrak{B}_2 y'z' + \mathfrak{B}_3 y'^2 = 0 \tag{91}$$

where the (x, y, z, p, q) in $\mathfrak{B}_1$, $\mathfrak{B}_2$, $\mathfrak{B}_3$ are the coördinates of an arbitrary surface element of C and $1:y':z'$ defines the direction element of C which is contained by that surface element.

Now, if in every surface element (x, y, z, p, q) of space there are three distinct directions which belong to curves that give rise to integrable fields, then (91) must have for given (x, y, z, p, q) three solutions for the ratio $y':z'$. Being a quadratic equation, (91) must therefore disappear identically, that is

$$\mathfrak{B}_1 = 0, \quad \mathfrak{B}_2 = 0, \quad \mathfrak{B}_3 = 0. \tag{92}$$

Adjoined to (88), these complete the five conditions (84) which are necessary and sufficient in order that ϵ be proportional to a contact transformation.

Since in the last discussion we have made use of only the first equation of (82), it is sufficient to require that each of the $5\infty^3$ manifolds specified give rise to an element-field having an integral surface consecutive to its base manifold.

PRINCETON UNIVERSITY,
PRINCETON, N. J.

A FIGURATRIX FOR DOUBLE INTEGRALS*

BY

PAUL R. RIDER

In a previous paper† the writer has called attention to a number of the interesting properties of the figuratrix in the calculus of variations. It is the purpose of the present paper to define a figuratrix for double integrals and to develop some of its important characteristics.

1. **A new form of the problem of double integrals.** The theory of extrema of double integrals has been treated for integrals of the form‡

$$\iint f(x, y, z, z_x, z_y)dxdy, \tag{1}$$

in which z and its partial derivatives are functions of x and y, and also for integrals of the form§

$$\iint F(x, y, z, x_u, y_u, z_u, x_v, y_v, z_v)dudv, \tag{2}$$

in which x, y, z are the coördinates of a surface

$$x = x(u, v), \quad y = y(u, v), \quad z = z(u, v). \tag{3}$$

Radon‖ has developed the theory for a third form of integral,

$$\iint F(x, y, z, A, B, C)dudv, \tag{4}$$

where the last three arguments of F are the usual determinants of surface theory,

---

* Presented to the Society, April 2 and September 9, 1926; received by the editors in July, 1926.

† *The figuratrix in the calculus of variations*, these Transactions, vol. 28 (1926), pp. 640–653. This paper will be referred to as Rider, Transactions.

‡ See, for example, Bolza, *Vorlesungen über Variationsrechnung*, chapter 13.

§ See Kobb, *Sur les maxima et les minima des intégrales doubles*, Acta Mathematica, vol. 16 (1892–93), pp. 65–140, vol. 17 (1893), pp. 321–43; and Kneser, *Lehrbuch der Variationsrechnung*, 1900, pp. 263–306.

‖ *Über einige Fragen betreffend die Theorie der Maxima und Minima mehrfachen Integrale*, Monatshefte für Mathematik und Physik, vol. 22 (1911), pp. 53–63. Radon considers a multiple integral, of which (4) is a special case. His notation is slightly different.

$$(5) \qquad A = \begin{vmatrix} y_u & z_u \\ y_v & z_v \end{vmatrix}, \quad B = \begin{vmatrix} z_u & x_u \\ z_v & x_v \end{vmatrix}, \quad C = \begin{vmatrix} x_u & y_u \\ x_v & y_v \end{vmatrix}.$$

It is found convenient in defining a figuratrix for double integrals to employ an integral of the form*

$$(6) \qquad \int\int f(x, y, z, \tau, \sigma) H du dv,$$

where, in the usual notation of the theory of surfaces,

$$(7) \qquad H = (EG - F^2)^{1/2} = (A^2 + B^2 + C^2)^{1/2}.$$

The quantities τ and σ are angles defined by the equations

$$(8) \qquad A = H \cos\tau \cos\sigma, \quad B = H \sin\tau \cos\sigma, \quad C = H \sin\sigma,$$

or

$$(9) \qquad \tau = \arctan(B/A), \quad \sigma = \arcsin(C/H).$$

The geometric meaning of τ and σ will be evident upon inspection of the accompanying figure, in which PN is normal to the surface (3) at the point P, and Pp, Pq, Pr are parallel to the coördinate axes Ox, Oy, Oz respectively.

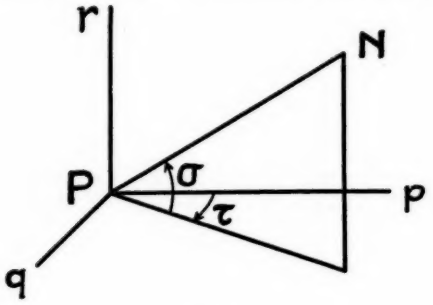

Fig. 1

The form (6) permits the use of the parametric representation of surfaces and avoids the homogeneity conditions required for integrals of form (2)† or form (4).‡ Moreover, it can be shown that any integral of

* Cf. Rider, *The space problem of the calculus of variations in terms of angle*, American Journal of Mathematics, vol. 39 (1917), pp. 241–56.

† See Kobb, loc. cit., p. 68.

‡ See Radon, loc. cit., p. 55.

the form (1), or any integral of the forms (2) or (4) satisfying the required homogeneity condition, can be reduced to the form (6).

2. **The Euler-Lagrange equations.** In order to develop the equations which must be satisfied by surfaces that minimize or maximize the integral (6), let us assume that the surface which affords the least or the greatest value is given by equations (3). If we take a neighboring surface

$$\bar{x} = x + \epsilon\xi, \quad \bar{y} = y + \epsilon\eta, \quad \bar{z} = z + \epsilon\zeta \tag{10}$$

and integrate the function fH over a region R of the uv-plane, we have

$$I(\epsilon) = \int_R \int f(\bar{x}, \bar{y}, \bar{z}, \bar{\tau}, \bar{\sigma})\bar{H}dudv, \tag{11}$$

where a bar over a variable indicates that its values on the surface (10) is to be used. For example,

$$\bar{\tau} = \arctan\frac{\bar{B}}{\bar{A}}, \quad \bar{A} = \begin{vmatrix} \bar{y}_u & \bar{z}_u \\ \bar{y}_v & \bar{z}_v \end{vmatrix}, \quad \bar{B} = \begin{vmatrix} \bar{z}_u & \bar{x}_u \\ \bar{z}_v & \bar{x}_v \end{vmatrix}.$$

Obviously $dI/d\epsilon$ must be equal to zero for $\epsilon=0$ if I is to be either a maximum or a minimum. We find that

$$\left.\frac{dI}{d\epsilon}\right|_{\epsilon=0} = \int_R \int \left[f \left.\frac{\partial \bar{H}}{\partial \epsilon}\right|_{\epsilon=0} + \left(f_x\xi + f_y\eta + f_z\zeta + f_\tau \left.\frac{\partial \bar{\tau}}{\partial \epsilon}\right|_{\epsilon=0} + f_\sigma \left.\frac{\partial \bar{\sigma}}{\partial \epsilon}\right|_{\epsilon=0} \right) H \right] dudv. \tag{12}$$

By differentiating the equation $\bar{H}^2 = \bar{A}^2 + \bar{B}^2 + \bar{C}^2$, we find that

$$\begin{aligned} \bar{H}\frac{\partial \bar{H}}{\partial \epsilon} &= \bar{A}\frac{\partial \bar{A}}{\partial \epsilon} + \bar{B}\frac{\partial \bar{B}}{\partial \epsilon} + \bar{C}\frac{\partial \bar{C}}{\partial \epsilon} \\ &= \bar{H}\left(\cos\bar{\tau}\cos\bar{\sigma}\frac{\partial \bar{A}}{\partial \epsilon} + \sin\bar{\tau}\cos\bar{\sigma}\frac{\partial \bar{B}}{\partial \epsilon} + \sin\bar{\sigma}\frac{\partial \bar{C}}{\partial \epsilon} \right). \end{aligned} \tag{13}$$

But

$$\left.\frac{\partial \bar{A}}{\partial \epsilon}\right|_{\epsilon=0} = \begin{vmatrix} \eta_u & z_u \\ \eta_v & z_v \end{vmatrix} + \begin{vmatrix} y_u & \zeta_u \\ y_v & \zeta_v \end{vmatrix},$$

and the values of $\partial\bar{B}/\partial\epsilon$ and $\partial\bar{C}/\partial\epsilon$ for $\epsilon=0$ can be obtained by cyclic permutation of x, y, z and ξ, η, ζ. Hence

$$\left.\frac{\partial \bar{H}}{\partial \epsilon}\right|_{\epsilon=0} = (y_v \sin\sigma - z_v \sin\tau\cos\sigma)\xi_u + (z_v \cos\tau\cos\sigma - x_v \sin\sigma)\eta_u$$

$$+ (x_v \sin\tau\cos\sigma - y_v \cos\tau\cos\sigma)\zeta_u$$

$$- (y_u \sin\sigma - z_u \sin\tau\cos\sigma)\xi_v - (z_u \cos\tau\cos\sigma - x_u \sin\sigma)\eta_v$$

$$- (x_u \sin\tau\cos\sigma - y_u \cos\tau\cos\sigma)\zeta_v. \tag{14}$$

Now we can readily find that

$$\left.\frac{\partial \bar{\tau}}{\partial \epsilon}\right|_{\epsilon=0} = \frac{1}{H}\left[- z_v \frac{\cos\tau}{\cos\sigma}\xi_u + z_v \frac{\sin\tau}{\cos\sigma}\eta_u + \left(x_v \frac{\cos\tau}{\cos\sigma} - y_v \frac{\sin\tau}{\cos\sigma}\right)\zeta_u \right.$$

$$\left. + z_u \frac{\cos\tau}{\cos\sigma}\xi_v - z_u \frac{\sin\tau}{\cos\sigma}\eta_v - \left(x_v \frac{\cos\tau}{\cos\sigma} - y_u \frac{\sin\tau}{\cos\sigma}\right)\zeta_v \right], \tag{15}$$

$$\left.\frac{\partial \bar{\sigma}}{\partial \epsilon}\right|_{\epsilon=0} = \frac{1}{H}[(y_v \cos\sigma + z_v \sin\tau\sin\sigma)\xi_u - (x_v \cos\sigma + z_v \cos\tau\sin\sigma)\eta_u$$

$$- (x_v \sin\tau\sin\sigma - y_v \cos\tau\sin\sigma)\zeta_u - (y_u \cos\sigma + z_u \sin\tau\sin\sigma)\xi_v$$

$$+ (x_u \cos\sigma + z_u \cos\tau\sin\sigma)\eta_v + (x_u \sin\tau\sin\sigma$$

$$- y_u \cos\tau\sin\sigma)\zeta_v]. \tag{16}$$

Substituting (14), (15), (16) in (12), and rearranging, we find that

$$\left.\frac{dI}{d\epsilon}\right|_{\epsilon=0} = \int_R \int [(f_x\xi + f_y\eta + f_z\zeta)H$$

$$+ (y_v r - z_v q)\xi_u + (z_v p - x_v r)\eta_u + (x_v q - y_v p)\zeta_u$$

$$- (y_u r - z_u q)\xi_v - (z_u p - x_u r)\eta_v - (x_u q - y_u p)\xi_v]\,du\,dv, \tag{17}$$

where

$$p = f\cos\tau\cos\sigma - f_\tau \sin\tau/\cos\sigma - f_\sigma \cos\tau\sin\sigma,$$

$$q = f\sin\tau\cos\sigma + f_\tau \cos\tau/\cos\sigma - f_\sigma \sin\tau\sin\sigma, \tag{18}$$

$$r = f\sin\sigma + f_\sigma \cos\sigma.$$

With the usual assumptions regarding the continuity of f and its derivatives, and with the further assumption that the functions ξ, η, ζ vanish on the contour of the region R, it is found by the ordinary methods that the following equations must be satisfied for a minimum or a maximum of the integral I:

$$f_x H - \frac{\partial(z, q)}{\partial(u, v)} + \frac{\partial(y, r)}{\partial(u, v)} = 0,$$

$$f_y H - \frac{\partial(x, r)}{\partial(u, v)} + \frac{\partial(z, p)}{\partial(u, v)} = 0, \tag{19}$$

$$f_z H - \frac{\partial(y, p)}{\partial(u, v)} + \frac{\partial(x, q)}{\partial(u, v)} = 0.$$

These equations are the so-called *Euler-Lagrange equations*.

3. **The figuratrix and its fundamental quantities.** We shall define the *figuratrix* of the integral (6), for the point $P(x, y, z)$ as origin, to be the surface defined by equations (18), in which p, q, r are rectangular coördinates of a point on the figuratrix, the arguments x, y, z of f and its derivatives being considered fixed in value. The parameters of this surface are consequently τ and σ.

The fundamental quantities of the figuratrix are*

$$\begin{aligned}
E &= (f_\tau \tan\sigma + f_{\tau\sigma})^2/\cos^2\sigma + (f + f_{\sigma\sigma})^2,\\
F &= (f_\tau \tan\sigma + f_{\tau\sigma})[(f + f_{\sigma\sigma}) + (f - f_\sigma \tan\sigma + f_{\tau\tau}/\cos^2\sigma)],\\
G &= (f_\tau \tan\sigma + f_{\tau\sigma})^2 + \cos^2\sigma(f - f_\sigma \tan\sigma + f_{\tau\tau}/\cos^2\sigma)^2,\\
D &= -(f\cos^2\sigma - f_\sigma \sin\sigma\cos\sigma + f_{\tau\tau}),\\
D' &= -(f_\tau \tan\sigma + f_{\tau\sigma}),\\
D'' &= -(f + f_{\sigma\sigma}),\\
H &= (EG - F^2)^{1/2} = \pm (DD'' - D'^2)/\cos\sigma.
\end{aligned} \tag{20}$$

These fundamental quantities play an important rôle in the theory of minimizing or maximizing the integral (6), as will be shown later.

4. **The transversality condition.** Let us note the relation between the integrands of integrals (6) and (4). We have

$$\begin{aligned}
&f(x,y,z,\tau,\sigma)H\\
&= f\left(x, y, z, \arctan\frac{B}{A}, \arcsin\frac{C}{(A^2 + B^2 + C^2)^{1/2}}\right)(A^2 + B^2 + C^2)^{1/2}\\
&= F(x,y,z,A,B,C).
\end{aligned} \tag{21}$$

By means of this relation we can transform many of the results of Radon into forms appropriate to the integral (6).

---

* See Rider, Transactions.

For example, the *transversality condition*,* which must be satisfied along the boundary of the extremal surface when this boundary is movable along another surface, readily assumes the form

$$p\cos\bar{\tau}\cos\bar{\sigma} + q\sin\bar{\tau}\cos\bar{\sigma} + r\sin\bar{\sigma} = 0, \tag{22}$$

where p, q, r are the functions (18) and refer to the extremal surface, and $\bar{\tau}$, $\bar{\sigma}$ give the normal direction of the boundary surface.

Let Q be the point (p, q, r) on the figuratrix for the point P. Let $PN(\tau, \sigma)$ be the normal to the extremal surface at the point P, and $PT(\bar{\tau}, \bar{\sigma})$ the normal to the boundary surface at this point. Then equation (22) states that $PT(\bar{\tau}, \bar{\sigma})$ *must be perpendicular to* PQ, *the radius vector of the figuratrix*, Q *being given by the parameter values* (τ, σ).

5. **A geometric interpretation of a certain necessary condition.** Radon† has stated that a certain quadratic form

$$\Phi_{ik}y_iy_k \tag{23}$$

cannot be negative upon an extremal surface if the integral (4) is a minimum, and cannot be positive if the integral is a maximum. In other words, the form (23) must be definite for an extremum. By making use of Radon's equation (11), we find that in the notation of the present paper,

$$\begin{aligned} H^3\Phi_{11} &= -Da_v^2 + 2D'a_vb_v - D''b_v^2, \\ H^3\Phi_{12} &= H^3\Phi_{21} = -Da_ua_v + D'(a_ub_v + a_vb_u) - D''b_ub_v, \\ H^3\Phi_{22} &= -Da_u^2 + 2D'a_ub_u - D''b_u, \end{aligned} \tag{24}$$

where

$$\begin{aligned} a_u &= z_u/\cos^2\sigma, \qquad b_u = z_u\tan\tau\tan\sigma - y_u/\cos\tau, \\ a_v &= z_v/\cos^2\sigma, \qquad b_v = z_v\tan\tau\tan\sigma - y_v/\cos\tau. \end{aligned} \tag{25}$$

It may be remarked that the Φ_{ik} can be expressed in a number of different ways, but for the purpose of establishing the results of this paper the forms given here are sufficient.

If the quadratic form (23) be definite it is necessary that $\Phi_{11}\Phi_{22} - \Phi_{12}^2$ be positive. We readily find that

---

* Radon, loc. cit., p. 58, equation (16).

† Loc. cit., p. 59.

$$\Phi_{11}\Phi_{22} - \Phi_{12}^2 = (a_u b_v - a_v b_u)^2 (DD'' - D'^2)/H^6 = \frac{DD'' - D'^2}{H^4 \cos^2 \sigma} = \frac{K}{H^2 \cos^2 \sigma}, \tag{26}$$

K being the total curvature* of the figuratrix. It follows that *the total curvature of the figuratrix must be positive at the point* (p, q, r) *given by the parameter values* (τ, σ) *which define the direction of the normal to an extremal surface.*

It should be noted that the Φ_{ik} correspond to F_1, F_2, F_3 in the notation of Kobb† and that Kobb's necessary condition $F_1F_2 - F_3^2 > 0$ is precisely the condition $\Phi_{11}\Phi_{22} - \Phi_{12}^2 > 0$.

6. **The *e*-function.** Let us suppose that we have a surface $\overline{S}$ lying in a field‡ of extremal surfaces. The *e*-function as defined by Radon§ is, in the notation employed in this paper,

$$\begin{aligned} E(x,y,z,A,B,C,\overline{A},\overline{B},\overline{C}) &= F(x,y,z,\overline{A},\overline{B},\overline{C}) - \overline{A}F_A(x,y,z,A,B,C) \\ &\quad - \overline{B}F_B(x,y,z,A,B,C) - \overline{C}F_C(x,y,z,A,B,C), \end{aligned} \tag{27}$$

where $\overline{A}$, $\overline{B}$, $\overline{C}$ define the direction of the normal to the surface $\overline{S}$ at the point (x, y, z), and A, B, C define the direction of the normal to the extremal surface through that point.

For an integral of the form (6), we find, by making use of the relation (21), that

$$\begin{aligned} E(x,y,z,\tau,\sigma,\bar{\tau},\bar{\sigma}) &= (\overline{A}^2 + \overline{B}^2 + \overline{C}^2)^{1/2}[f(x,y,z,\bar{\tau},\bar{\sigma}) \\ &\quad - (p \cos \bar{\tau} \cos \bar{\sigma} + q \sin \bar{\tau} \cos \bar{\sigma} + r \sin \bar{\sigma})], \end{aligned} \tag{28}$$

where $\bar{\tau}$, $\bar{\sigma}$ refer to the surface $\overline{S}$ and τ, σ to the extremal surface. We shall discard the positive factor $(\overline{A}^2+\overline{B}^2+\overline{C}^2)^{1/2}$ and define

$$e(\tau,\sigma,\bar{\tau},\bar{\sigma}) = f(\bar{\tau},\bar{\sigma}) - (p \cos \bar{\tau} \cos \bar{\sigma} + q \sin \bar{\tau} \cos \bar{\sigma} + r \sin \bar{\sigma}), \tag{29}$$

omitting, for the sake of simplicity, the arguments x, y, z.

---

* See Eisenhart, *Differential Geometry*, p. 123.

† Loc. cit.

‡ For the notion of a field of extremal surfaces, see Bolza, loc. cit., pp. 684 ff.

§ Loc. cit., p. 60.

Let now PQ be the radius vector of the figuratrix for the point P, and PS the projection of PQ upon the line passing through P in the direction $(\bar{\tau}, \bar{\sigma})$. If $P\overline{N}=f(\bar{\tau}, \bar{\sigma})$ be marked off on the line having the direction $(\bar{\tau}, \bar{\sigma})$, then *the e-function is the line-segment SN.* This can be shown precisely as in the author's paper on the figuratrix for simple integrals,* as can also the following statement:

The e-function is the distance from the point $\overline{Q}(\bar{\tau}, \bar{\sigma})$ on the figuratrix to the tangent plane at the point $Q(\tau, \sigma)$, except for an infinitesimal of at least the third order with respect to $(\bar{\tau}-\tau)$ and $(\bar{\sigma}-\sigma)$.

---

* Rider, Transactions.

WASHINGTON UNIVERSITY,
ST. LOUIS, MO.

MANIFOLDS WITH A BOUNDARY AND THEIR TRANSFORMATIONS*

BY

SOLOMON LEFSCHETZ

This is the continuation of the paper which appeared in the January, 1926, number of these Transactions.† Its chief object is to extend to an M_n with a boundary the results already obtained for transformations of manifolds without boundary. To these already treated in full we devote a few pages chiefly to elucidate and simplify certain points of importance for the extension. We have succeeded in deriving coincidence and fixed points formulas for the two types of transformations that are alone amenable to anything like a general treatment and extended the formulas of this and the preceding paper to transformations between two different manifolds with or without a boundary. As an incidental acquisition there should be pointed out some highly interesting topological propositions obtained in Parts II, III. Of importance also is the fact that by means of ample use of matrices we have been able to put all coincidence formulas of this and the previous paper in very simple and manageable form.

I. General remarks on manifolds

1. **A theorem on intersecting complexes.** Let C_h, C_k, M_p be two complexes and a manifold on M_n, all polyhedral and with C_k a sub-complex of M_p. We wish to prove that *the intersections $C_h \cdot C_k$ taken on M_n and $(C_h \cdot M_p) \cdot C_k$ taken on M_p coincide and when $h+k=n$ the related Kronecker indices are equal.* We assume that as regards all intersections to be considered the restrictions of Tr., No. 15, are fulfilled. The problem is then reduced at once to the case where the complexes are simplexes and the manifolds

---

* Presented to the Society, October 30, 1926; received by the editors in November, 1926. See also Proceedings of the National Academy of Sciences, vol. 12 (1926), p. 737.

† Referred to in the sequel as Tr. Unless otherwise stated, the notations, terminology, assumptions, etc., of that paper will apply directly here. The only changes will be actually in the definition of an M_n, and using for cycles besides Γ, also γ as in Tr., Part II, and later other letters for special types. Since we shall make considerable use of matrices we may as well give our notations here. All our matrices will have integer terms. Any matrix will be designated as its generic element with position indices omitted. The transverse of a matrix m will be called m'; when m is square its determinant is denoted by $|m|$ and the sum of the terms in its principal diagonal is called its *trace*.

linear spaces. The natural procedure is as in Tr., Part I, §2, to compare indicatrices. The indicatrix of Tr., No. 7, is now so chosen that $A_0A_1\cdots A_q$, $q=p+h-n$, be on $C_h\cdot M_p$, $A_0A_1\cdots A_qA_{h+1}\cdots A_r$, $r=p+q-n$, be on M_p and that multiplied respectively by α_q, α_p, they constitute indicatrices of their complexes. Let also $\alpha_l' A_0A_1\cdots A_l$ be an indicatrix of $(C_h\cdot M_p)\cdot C_k$ when the intersection is taken on M_p. Then the relations

$$\alpha_h\alpha_k\alpha_l\alpha_n = \alpha_h\alpha_p\alpha_q\alpha_n = \alpha_q\alpha_k\alpha_l'\alpha_p = 1 \tag{1.1}$$

define the orientations of $C_l=C_h\cdot C_k$ and $C_h\cdot M_n$ taken on M_n, and that of $(C_h\cdot M_p)\cdot C_k$ taken on M_p. Since the α's are all ±1, from (1.1) follows at once

$$\alpha_k\alpha_l' = \alpha_p\alpha_q = \alpha_h\alpha_n = \alpha_k\alpha_l, \tag{1.2}$$

therefore

$$\alpha_l' = \alpha_l. \tag{1.3}$$

The cells of the complexes to be compared are the same, and by (1.3) similarly oriented. Therefore

$$C_h\cdot C_k = (C_h\cdot M_p)\cdot C_k, \tag{1.4}$$

as we wished to prove. As we know from Tr., No. 8, when $l=0$, we merely have a Kronecker index to consider and then

$$(C_h\cdot C_{n-h}) = ((C_h\cdot M_p)\cdot C_{n-h}), \tag{1.5}$$

which may also be established directly as above by comparison of indicatrices.

Of particular interest is the case when the C's are cycles. The passage through suitable approximations, as in Tr., Part I, §4, will enable us to drop all restrictions as to them provided their intersection does not meet the boundary of M_p. For the latter being polyhedral, the approximations can always be so carried out that C_l remains on it.

2. In the applications that we have especially in view, $p=n-1$, $k=n-h$, so that we deal with a Kronecker index. Furthermore M_{n-1} will there be a subcomplex of the defining C_n of M_n and the preceding result does not apply outright. The extension is, however, easy on this basis: C_h and C_{n-h} are assumed as general as possible and hence their intersections are isolated points, of which any one, say A, is on an E_{n-1} of C_n. As far as the contribution of A to the index is concerned, only the two n-cells E_n, E_n' of C_n incident with E_{n-1} are involved. The situation is then the same as if M_n were reduced to $E_n+E_n'+$their boundaries. But this system is obviously

homeomorphic with preservation of structure* to a similar one in S_n. Practically this means that we may assume that M_n is an S_n. On the latter, however, we can always choose a defining C_n with A on an n-cell of it. Then M_p will not be a subcomplex of it, or rather its cells through A will not be cells of C_n. Therefore we are back to the case already considered and the conclusion is the same.

3. **On the Kronecker index.** Until further notice we assume that M_n is *without boundary*. In an important paper Veblen has had occasion to define so-called *intersection numbers* for associated complexes of special type on M_n.† Since we have applied some of his results it is important to show that his numbers are merely the Kronecker indices of the complexes.

We begin with Poincaré congruences

$$C_k \equiv \Gamma_{k-1}, \qquad C_{n-k+1} \equiv \Gamma_{n-k}, \tag{3.1}$$

where the cycles have no common points. They give rise as in Tr., No. 18, to the relation

$$(C_k \cdot \Gamma_{n-k}) = (-1)^k (\Gamma_{k-1} \cdot C_{n-k+1}), \tag{3.2}$$

valid without any other restrictions than the one just stated concerning the cycles. That follows at once by passing to polyhedral approximations for which (3.2) holds. Since the indices are defined in each case by means of the approximations the relation is valid for the initial complexes.

---

* Two sets of simplicial cells, $\{e\}$, $\{e'\}$, are said to have the same structure, whenever to each e there corresponds one and only one e' of same dimensionality and conversely, and when furthermore the incidence relations between any two e's and the corresponding e''s are the same.

† These Transactions, vol. 25 (1923), pp. 540–550. Substantially the same results, derived in similar fashion, were also obtained simultaneously but independently by Hermann Weyl, Revista de Matematica Hispaño Americana, 1923. Regarding the index, I recently received a communication from Weyl in which he points out that, unknown to me, he had proved its independence from the defining C_n: (a) for $n=2$ in Jahresbericht der Deutschen Mathematiker-Vereinigung, vol. 25 (1916), p. 225, also Note to the second edition of *Die Idee der Riemannschen Flächen*; (b) for $(\Gamma_1 \cdot \Gamma_{n-1})$ and any n in the Revista paper. In the same paper he also points out that for $\mu = \frac{1}{2}n$ even, it may not be possible to have a canonical set whose matrix of indices is the identity. The bearing on the coincidence formulas in Tr. is that one must have two associated sets as when $\mu \neq n/2$, with T operating on one and T' on the other. Let these sets be γ^i, γ'^i, with $\gamma'^i \approx \Sigma g_{ij}\gamma^j$. All matrices g corresponding to such a pair of associated sets are of the form pgp', where p is an arbitrary square matrix of order $R_{n/2}$ with $|p| = \pm 1$. The properties of the whole class of such matrices are invariants of M_n, another form of a remark made by Weyl, loc. cit.

In a letter received in early December, Weyl communicated to me substantially the same derivation as mine of the matrix formulas (10.3) and (39.1) of this paper from those of Tr., Part II (see also the note in the Proceedings of the National Academy of Sciences, for December, 1926). He has thus confirmed my results, at an important point.

4. Let us now adopt the notations of Tr., No. 23. We designate furthermore by $\overline{C}_n$ the dual of C_n, whose cells are subcomplexes of C_n' also. Any one $\overline{E}_{n-k}$ is a sum of cells of C_n' of type $(k, p_1, \cdots, p_i)$, with a common vertex A^k, and the p's all $>k$. The $(n-k)$-cells of the sum are of type $(k, k+1, \cdots, n)$. A similar statement holds for the cell E_k of C_n that carries A^k, except that now the types are $(q_1, \cdots, q_j, k)$, $q_i<k$ (Coll. Lect.,* p. 89).

Let A^{k-1} be any vertex on a cell E_{k-1} of the boundary of E_k. It is on a cell $\overline{E}_{n-k+1}$ of $\overline{C}_n$, which as before is the sum of certain cells of C_n' with A^{k-1} for vertex. We assume E_{k-1} positively related to E_k, hence a positive cell of its boundary Γ_{k-1}, and $\overline{E}_{n-k+1}$ so sensed that $\overline{E}_{n-k}$ is a positive cell of its boundary Γ_{n-k}. Now the two cycles have no common points, since these would have to be vertices of C_n', while the vertices they carry are of the incompatible types A^{k+i}, A^{k-i}, $i\neq 0$. Hence (3.2) is applicable here and

$$(E_k \cdot \overline{\Gamma}_{n-k}) = (-1)^k(\Gamma_{k-1} \cdot \overline{E}_{n-k+1}). \tag{4.1}$$

E_k and $\overline{\Gamma}_{n-k}$ meet at A^k, and nowhere else. For all cells of C_n' of which the cycle is made up are of type $(k, p_1, \cdots, p_r)$, $p_i>k$. Therefore (Tr., No. 23), they are on cells of k dimensions of C_n, unless they are merely vertices and of the same type as A^k. In that case they are on k-cells with one and only one such vertex on each k-cell. Hence the k-cell E_k of C_n can only meet $\overline{\Gamma}_{n-k}$ at a single point which can only be A^k. Therefore

$$(E_k \cdot \overline{\Gamma}_{n-k}) = (E_k \cdot \overline{E}_{n-k}). \tag{4.2}$$

Due to the symmetrical relation of C_n' to the dual complexes, it is not necessary to repeat the discussion for the second index in (4.1) and we infer at once

$$(\Gamma_{k-1} \cdot \overline{E}_{n-k+1}) = (E_{k-1} \cdot \overline{E}_{n-k+1}). \tag{4.3}$$

Therefore in place of (2.1)

$$(E_k \cdot \overline{E}_{n-k}) = (-1)^k(E_{k-1} \cdot \overline{E}_{n-k+1}). \tag{4.4}$$

Now this is precisely the relation proved by Veblen for his intersection numbers in §4 of his paper. Therefore in proving that they are merely Kronecker indices, say for a given k, the latter may be increased or decreased by one unit. Hence we may assume $k=n$. But for this special value Veblen's definition reduces essentially to ours. *Therefore his numbers are Kronecker indices for every k.*

---

* *Analysis Situs*, by Oswald Veblen (The Cambridge Colloquium, Part II), will be referred to as Coll. Lect. throughout the present paper.

5. We are then justified in taking over Veblen's theorems bodily. Of particular significance are these two:

I. *In order that* $\gamma_\mu \approx 0$ *on* M_n *without boundary (i.e., that it be a zero-divisor or bounding cycle) it is necessary and sufficient that* $(\gamma_\mu \cdot \gamma_{n-\mu}) = 0$ *whatever* $\gamma_{n-\mu}$.

II. *Let* $\gamma_\mu^i, \gamma_{n-\mu}^j$ *run respectively through the elements of two fundamental sets for their dimensions. Then the rank of*

$$\| (\gamma_\mu^i \cdot \gamma_{n-\mu}^j) \| \tag{5.1}$$

is R_μ, *and its invariant factors are all unity.*

These propositions are proved by Veblen only when in each pair one cycle is a subcomplex of C_n, the other one of the dual $\overline{C}_n$. As every cycle is homologous to a subcomplex of $\overline{C}_n$ (Tr., No. 27, Remark), and as the index is invariant with respect to homology, the two propositions are true without restrictions.

In our previous paper we actually made use only of the first part of II, that is, of the fact that (5.1) is of rank R_μ, and proved the other part directly. Of I also, the necessary condition alone was needed and established directly in Part I. The only question that could have been raised is then as to whether the R's in the formulas are the connectivity numbers and not merely the ranks of the matrices (5.1) and we have just answered it in the negative.

6. Let us return to the fundamental set $\gamma_\mu^1, \gamma_\mu^2, \cdots, \gamma_\mu^{R_\mu}$ relative to the operation $\approx$ as considered in Tr., p. 37. It has the property that any γ_μ of M_n is a combination of the cycles of the set plus a zero-divisor. Hence a fundamental set as to $\sim$ is obtained by merely adding zero-divisors to the γ's. For example, the set in Coll. Lect., p. 117, is of this very nature, and so are the canonical sets of our first paper, but we need not limit ourselves to these.

Let $\gamma_{n-\mu}^1, \gamma_{n-\mu}^2, \cdots, \gamma_{n-\mu}^{R_\mu}$ correspond in analogous fashion to the dimensionality $n-\mu$. As is well known the number of the cycles is again R_μ (Poincaré). Complete the two sets by zero-divisors so as to have fundamental sets for $\sim$, then form (5.1). By §3, the matrix will merely consist of the square array

$$L_\mu = \| (\gamma_\mu^i \cdot \gamma_{n-\mu}^j) \| \qquad (i,j = 1,2, \cdots, R_\mu) \tag{6.1}$$

bordered with rows and columns of zeros. Hence the determinant $|L_\mu| = \pm 1$, for in absolute value it is the product of the invariant factors of L_μ, all equal to one.

7. **Continuous transformations.** First observation of very general nature: All of Tr. Part I, goes through whether the cells of C_n are simplicial or merely convex (i.e., corresponding to convex polyhedral regions on the representative polyhedron Π_n). The bearing of this becomes clear when we remember that even when the cells of C_p, C_q are simplicial, those of $C_p \times C_q$ are merely convex. Hence it is proper to apply to a product manifold all the approximation work (loc. cit.). Of course the more general type of C_n, practically the same as Veblen's, possesses a regular subdivision of the restricted type (Coll. Lect., p. 85).

This important point settled let us return to the situation of Tr., Part II, §2, particularly as regards the definition of $T\gamma_\mu = \bar{\gamma}_\mu$. It is obtained by means of an approximation $\bar{\Gamma}_n$ to Γ_n, defining cycle of the transformation T. The question arises, however, whether $T\gamma_\mu$ is unique. To show that such is the case let a second defining complex C_n^0 lead to $\Gamma_n^0, \gamma_\mu^0, \bar{\gamma}_\mu^0$. We do not exclude the possibility that C_n^0 coincides with C_n. Then by Tr., Part I, §6, if the approximations are sufficiently close,

$$\bar{\Gamma}_n^0 \cdot \gamma_\mu^0 \times M_n' \sim \bar{\Gamma}_n \cdot \gamma_\mu \times M_n' \qquad (\text{mod } M_n \times M_n'). \tag{7.1}$$

Let the difference of the two sides bound $C_{\mu+1}$. When the point $A \times B$ describes it, B describes on M_n' a singular image $C_{\mu+1}'$ of the complex, and when $A \times B$ describes the boundary of $C_{\mu+1}$, B describes $\bar{\delta}_\mu - \bar{\delta}_\mu^0$. Hence this last cycle bounds on M_n', and $\bar{\gamma}_\mu - \bar{\gamma}_\mu^0$ therefore bounds on M_n, that is $\bar{\gamma}_\mu \sim \bar{\gamma}_\mu^0$ (mod M_n), which proves the uniqueness of $T\gamma_\mu$.

8. Throughout our first paper we have assumed that an oriented zero-cell is considered as a cycle of zero dimensions. (See in particular No. 63.) This justifies, for example, the theorem of No. 52, our assertion as to the Euler characteristic, No. 71, etc. Veblen in Coll. Lect., p. 110, does not consider such cycles. The advantage of our procedure is readily perceived. Two points similarly oriented on the same connected piece of a C_n constitute homologous cycles, for their difference is the boundary of an obvious one-cell. Hence R_0 is the number of distinct connected pieces of C_n, and if it is an M_n without boundary, $R_0 = R_n$, and Poincaré's duality formula holds without exception. His formula for the Euler characteristic also takes the simpler form, assumed in Tr., No. 71,

$$\sum(-1)^i \alpha_i = \sum(-1)^i R_i. \tag{8.1}$$

The convention of Tr., No. 58, reduces to the following for $\mu = 0$: The transform of $\gamma_0 = A$, point of M_n, is $\alpha_0 A$, where α_0 is as in Tr., No. 56.

Incidentally the orientation of Γ_n, defined (loc. cit.) by $\alpha_0 \geqq 0$, is indeterminate when $\alpha_0 = 0$. In that case we orient the cycle arbitrarily. This has no great importance, as a change of orientation merely changes the sign of certain Kronecker indices, whose absolute value, however, is alone of interest.

What matters chiefly is to make sure that the fundamental formula (59.1) holds without exception. The proof goes through in fact even for $\mu = 0$, n, but the direct verification for both cases is very simple. The case $\mu = n$ has already been considered in No. 63, but there is a simpler and more direct verification, as we shall presently see.

Let first $\mu = 0$. Denote again by θ, as in Tr., No. 61, the contribution of a certain point $A \times B$ to α_0. With the same notations as there used, except that the cycle is $\overline{\Gamma}_n$, if it carries the cell $E_n = A \times B \cdots A_n \times B_n$ then θE_n is its indicatrix as cell of $\overline{\Gamma}_n$. It follows $\alpha_0 = \sum\theta$, where the sum is extended to all points $A \times B$ of $\overline{\Gamma}_n$ corresponding to a fixed A. If $\overline{T}$ is the transformation defined by $\overline{\Gamma}_n$, we may think of them as the points $A \times B$ whose B is the image of some $\overline{T}A$. The verification of (59.1) for $\mu = 0$ requires that

$$(8.2) \quad (\Gamma_n \cdot A \times M_n') = (\overline{\Gamma}_n \cdot A \times M_n') = (\overline{T}A \cdot M_n') = \alpha_0 (A \cdot M_n') = \alpha_0,$$

which is in accordance with the formula in Tr., No. 56.

Let now $\mu = n$, and denote by ϵ the same integer as in Tr., No. 61; $\epsilon B B_1 \cdots B_n$ is then the indicatrix of M_n' at B and ϵ has the sign of $|Y_n|$. Hence as at the end of No. 61, $A \times B$ contributes $(-1)^n \epsilon\theta$ to $(\overline{\Gamma}_n \cdot M_n \times B)$ and therefore

$$(8.3) \qquad (\overline{\Gamma}_n \cdot M_n \times B) = (-1)^n \sum \epsilon\theta.$$

Here the sum is extended to all points of Γ_n with a fixed B, or to all points $A \times B$ of the cycle such that among the points $\overline{T}A$ there is one whose M_n' image is B.

Now the image of $\overline{T}C_n$ is a polyhedral γ_n on C_n. We may so subdivide C_n, say into C_n^0, that $\overline{T}C_n$ be a subcomplex of the subdivision. Then any particular cell E_n of C_n^0 will count say k' times positively and k'' times negatively among the cells of TC_n, and we shall have from the above, if B is on E_n,

$$(8.4) \qquad k' - k'' = \sum\theta\epsilon = \alpha_n ;$$

$$(8.5) \qquad \overline{T}M_n = \alpha_n \cdot M_n.$$

The verification of (8.2) for $\mu = n$ requires here that

$$(8.6) \qquad (\bar{\Gamma}_n \cdot M_n \times B) = (\bar{\Gamma}_n \cdot M_n \times B) = (-1)^n(\bar{T}M_n \cdot A)$$
$$= (-1)^n\alpha_n(M_n \cdot A) = (-1)^n\alpha_n,$$

which comparison with (8.3) shows is correct.

9. **Invariant form for coincidence and fixed point formulas.** The formulas given in Tr. are rather involved and furthermore depend upon a special choice of fundamental sets. It so happens that by making use of matrices there can be derived formulas independent of the particular fundamental sets relative to the operation $\approx$ that may be chosen. To begin with, the first formula on p. 43, Tr., reads*

$$(9.1) \qquad (-1)^{\mu(n+1)}L_\mu\epsilon_{n-\mu}L_\mu = \alpha_\mu L_\mu,$$

where L_μ is as defined in §6 of the present paper. Since its determinant is not zero, this gives at once

$$(9.2) \qquad (-1)^{\mu(n+1)}L_\mu\epsilon_{n-\mu} = \alpha_\mu,$$

and then

$$(9.3) \qquad \epsilon_{n-\mu} = (-1)^{\mu(n+1)}L_\mu^{-1}\alpha_\mu.$$

This solves then the problem of expressing the ϵ's in terms of the α's for any choice of fundamental sets as to $\approx$, and not merely for the canonical sets. The explicit formulas of No. 64, Tr., for canonical sets follow from these simply by replacing the L's by the form corresponding to each μ.

10. For the coincidence formula the starting point is as in Tr., No. 70, the relation independent of the choice of sets

$$(10.1) \qquad (\Gamma_n \cdot \Gamma_n') = \sum \epsilon_{n-\mu}^{ij}\eta_\mu^{hk}(\gamma_{n-\mu}^i \times \delta_\mu^j \cdot \gamma_\mu^h \times \delta_{n-\mu}^k).$$

By applying Tr., (53.3), and then replacing afterwards the δ's by γ's, this becomes

$$(10.2) \qquad (\Gamma_n \cdot \Gamma_n') = \sum(-1)^\mu\epsilon_{n-\mu}^{ij}\eta_\mu^{hk}(\gamma_{n-\mu}^i \cdot \gamma_\mu^h)(\gamma_\mu^j \cdot \gamma_{n-\mu}^k)$$
$$= \sum(-1)^\mu \text{ trace } \epsilon_{n-\mu}'L_{n-\mu}\eta_\mu L_\mu'.$$

By (9.2) or (9.3) applied to both transformations, and recalling that $(ab)' = b'a'$, we have

$$(10.3) \qquad \epsilon_{n-\mu}'L_{n-\mu}\eta_\mu L_\mu' = (L_\mu^{-1}\alpha_\mu)'\beta_{n-\mu}L_\mu'$$
$$= (L_\mu\beta_{n-\mu}'L_\mu^{-1}\alpha_\mu)'.$$

---

* In that formula $\gamma_{n-\mu}^j$ must be replaced by $\gamma_{n-\mu}^\nu$.

Since transposed matrices have equal traces, (10.2) may be replaced by

$$(10.4) \qquad (\Gamma_n \cdot \Gamma_n') = \sum(-1)^\mu \text{ trace } L_\mu \beta'_{n-\mu} L_\mu^{-1} \alpha_\mu .$$

This is the final form that we need for the coincidence formula. It is manifestly more compact and clear than what is found in No. 70, Tr.,* and has also the requisite invariant form most appropriate for manifolds with boundary. For the fixed point formula we choose here the second transformation as the identity. Then the β's are all unit matrices and (10.4) reduces to

$$(10.5) \qquad (\Gamma_n \cdot \Gamma_n^0) = \sum(-1)^\mu \text{ trace } \alpha_\mu .$$

This is (71.1), Tr., with the two cycles interchanged. However, with the situation chosen there, μ should be replaced by $n-\mu$ in (71.1), or, what is equivalent, the two cycles interchanged on the left. Here again, as in loc. cit., for a transformation of the same class as the identity, or more generally for one which merely adds zero-divisors to any cycle,

$$(10.6) \qquad (\Gamma_n^0 \cdot \Gamma_n^0) = \sum(-1)^\mu R_\mu ,$$

the Euler characteristic.

11. While dealing with transformations, let us bring out the following interesting property: *The transform of a zero-divisor or cycle is also a zero-divisor or cycle.* In signs, if $\gamma_\mu \approx 0$, also $\bar{\gamma}_\mu = T\gamma_\mu \approx 0$. For then $\gamma_\mu \times \delta_{n-\mu}$ is also a zero-divisor or cycle for $M_n \times M_n'$, hence, by §3 and Tr. (59.1),

$$(11.1) \qquad 0 = (\Gamma_n \cdot \gamma_\mu \times \delta_{n-\mu}) = (-1)^\mu (\bar{\gamma}_\mu \cdot \gamma_{n-\mu})$$

for every $\gamma_{n-\mu}$, from which at once $\bar{\gamma}_\mu \approx 0$.

II. Manifolds with a boundary

12. We propose to modify somewhat the definition of manifolds of our earlier paper. The difference, however, pertains only to the boundary and since it has played no direct part there, all results so far obtained will continue to hold. Let C_n be a complex. Consider the star of cells whose center is a given E_k of C_n. Between its cells there take place the same incidence relations as between the elements of a certain C_{n-k-1}: the h-cells of the latter correspond to the $(h+k+1)$-cells of C_n incident with E_k, or the star of cells of center E_k. Now C_n defines a manifold M_n when C_{n-k-1} is homeomorphic

---

* In the first formula of No. 70 replace $\sim$ by $\times$.

(a) with the boundary of an E_{n-k} (sphere of S_{n-k});

or

(b) with an E_{n-k-1} plus its boundary.

E_k is an *interior* or a *boundary* cell of M_n according as (a) or (b) is fulfilled. The set F_{n-1} of all boundary cells is the boundary of M_n.

Due to (b), the C_{n-k-2} image of the star of F_{n-1} with center E_k assumed on F_{n-1} is homeomorphic with the boundary of an E_{n-k-1}. Hence (a) holds for every cell of F_{n-1}, which is thus, itself, an M_{n-1} without boundary.

The manifold conditions may be replaced by others equivalent and often more convenient. Instead of considering the h-cells of the star of center E_k as $(h-k-1)$-cells of a certain complex, let us think of them as $(h-k)$-cells of a new system s_{n-k} with a unique zero-cell corresponding to E_k. Then in place of (a) and (b) we may obviously impose the following conditions:

(a′) when E_k is an interior cell, s_{n-k} is homeomorphic with an E_{n-k};

(b′) when E_k is on the boundary, s_{n-k} is homeomorphic with a star of cells of S_{n-k} that constitutes an E_{n-k} plus an E_{n-k-1} on its boundary, or what is the same thing, with a hemispherical region of S_{n-k} plus its flat base.*

The conditions here imposed for an M_n are more stringent for the boundary than Veblen's (Coll. Lect., p. 88). They are, however, in the nature of a certain homogeneity requirement along the boundary and entirely similar to what is imposed on the interior. It will also be observed that they do not demand that the cells of C_n be *simplicial*, but merely that they be *convex*.

13. *When C_n satisfies the manifold conditions so does any subdivision of it, C_n'.* This is proved in outline as follows. With each flat E_k on C_n we associate a system such as $\{e\}$ of Tr., No. 3, where e represents a class of incident $(k+1)$-cells on the same half S_{k+1}. The system $\{e\}$ is the same for two flat k-cells with a k-subcell in common. The $(h+k+1)$-cells incident with E_k may be considered as h-cells made up with the similar $(k+1)$-cells as points, that is with the e's as points. Hence when E_k is a cell of any particular subdivision C_n' of C_n, $\{e\}$ is homeomorphic with the C'_{n-k-1} similar to C_{n-k-1} of §12. It is then sufficient to show that when C_n behaves as desired, every $\{e\}$ obeys conditions (a), (b).

Let then E_k carry E_{k-1}, with a corresponding system $\{e'\}$. When $\{e\}$ behaves as desired so does $\{e'\}$. For to each e there corresponds a one-cell

* One is tempted to replace (b′) by the simpler "s_{n-k} is an E_{n-k} plus an E_{n-k-1} on its boundary." Unfortunately to show that this is equivalent to (b′) we need the following theorem: Two h-cells can be homeomorphically transformed into one another in such manner that two $(h-1)$-cells of their boundaries are similarly transformed. For $h=2$ this goes back to the Jordan curve theorem, but beyond that there is no proof.

of e''s with fixed end points (images of E_k itself). Hence the relation between the two systems is like that between a sphere or hemisphere in S_{n-k} (a sphere when E_k is an interior cell, a hemisphere otherwise) and its locus when the space, immersed in an S_{n-k+1}, rotates through an angle π around a diameter. It follows that for our purpose $\{e\}$ may replace $\{e'\}$ and E_k replace E_{k-1}. Ultimately, then, we shall merely have to consider some cell of C_n itself; that is, as asserted, the correct behavior of $\{e\}$ follows from that of every C_{n-k-1} attached to the cells of C_n.

14. From (a) and (b) as applied to $(n-1)$-cells it follows that every interior $(n-1)$-cell of C_n separates two n-cells, and every boundary $(n-1)$-cell is on a unique n-cell of C_n; F_{n-1} is then the sum of all $(n-1)$-cells on a unique n-cell of C_n. *We assume again that M_n is orientable.* Then F_{n-1} will also be orientable. For let us orient C_n and then sense each E_{n-1} of F_{n-1} positively in relation to the E_n that it bounds. Between the n- and $(n-1)$-cells incident with a given E_{n-2} of F_{n-1} we can write down the same Poincaré congruences as for the one- and zero-cells of a polygonal line. Hence the two end $(n-1)$-cells, which are those of E_{n-1} incident with E_{n-2}, are oppositely related to E_{n-2}, and F_{n-1} is oriented. Its orientation as thus fixed shall be preserved throughout. It corresponds to the congruence $M_n \equiv F_{n-1}$.

15. **The auxiliary manifold V_n.** We assume henceforth that M_n has a boundary F_{n-1}. Take, then, another copy $\overline{M}_n$ of the manifold and piece the two together along corresponding boundary points. *The new configuration V_n so obtained is an M_n without boundary.* (Any element of $\overline{M}_n$ corresponding to a given one of M_n will be called its conjugate and designated by the same letter barred.) If C_n is the basic defining complex of M_n, we use $\overline{C}_n$ for $\overline{M}_n$ and $C_n+\overline{C}_n$ for V_n. Then if E_k is the cell of M_n in §12, when it is not on F_{n-1}, the complex C_{n-k-1} plays the same part for it relative to V_n as to M_n. Hence it behaves then according to (a), and similarly for $\overline{E}_k$ and $\overline{C}_{n-k-1}$. However, when E_k is on F_{n-1}, in place of C_{n-k-1} we have $C_{n-k-1}+\overline{C}_{n-k-1}$. As this set is composed of two $(n-k-1)$-cells pieced together along their boundaries, it is homeomorphic to the boundary of an E_{n-k}. This is seen at once by referring to the piecing together of two hemispheres in S_{n-k} into a sphere of that space. Hence E_{n-k} behaves again in accordance with (a), which proves our assertion as to V_n.

If E_n is an n-cell of C_n, E_n' its indicatrix, we sense $\overline{E}_n$ of $\overline{M}_n$ by $-\overline{E}_n'$, hence $V_n = M_n - \overline{M}_n$. The importance of V_n is due to the fact that the solution of the coincidence problem for pairs of transformations of M_n will be reduced to the same problem for pairs of associated transformations of V_n.*

* This or a similar procedure has been followed by other authors dealing with this question. See, for example, Brouwer, Comptes Rendus, vol. 168 (1919), p. 1042; Alexander, these Transactions, vol. 23 (1922), pp. 89–95.

It is evident that its topological properties are really inherent properties of M_n itself. We shall be particularly concerned with the study and disposition of its fundamental sets.

16. **Fundamental sets for V_n.** As several distinct types of cycles will have to be considered we shall avoid excess of indices by using not only Γ, γ but also G, Δ, D to designate them.

Let $\Gamma_\mu^1, \cdots, \Gamma_\mu^p$ be a fundamental set for the μ cycles of M_n, and consider all possible homologies

$$\sum t_i \Gamma_\mu^i \sim \text{a cycle of } F_{n-1} \pmod{M_n}. \tag{16.1}$$

By paraphrasing a well known process* we may readily establish that these homologies are sums of multiples of a finite number of the same type which constitute a fundamental set for them. The members of the fundamental set and also the cycles can then be combined in such a fashion as to have a new fundamental set of Γ's (for which we keep the same designation as above) with fundamental homologies

$$\theta_i \Gamma_\mu^{r_\mu^1+i} \sim \text{a cycle of } F_{n-1} \pmod{M_n} \qquad (i = 1,2, \cdots, p - r_\mu^1). \tag{16.2}$$

The operations referred to correspond to elementary transformations on the matrix of the coefficients of the fundamental homologies. The first r_μ^1 cycles are not related by any homology such as (16.1) and in particular they are entirely independent; the remaining cycles have some non-zero multiple homologous to a cycle on the boundary. Between the cycles $\Gamma_\mu^{r_\mu^1+i}$ there may exist homologies mod M_n. Reducing those as above, we shall replace the cycles by a new set G_μ^j, $j=1, 2, \cdots, p-r_\mu^1$, whose first say r_μ^2 elements are independent while the others are ≈ 0. Of course $r_\mu^1 + r_\mu^2 = R_\mu$, the μth connectivity index of M_n. The cycles $\Gamma_\mu^i, i=1, 2, \cdots, r_\mu^1, G_\mu^j, j=1, 2, \cdots, r_\mu^2$, constitute a fundamental set for M_n relative to the operation $\approx$.

It will be convenient to call a cycle *symmetric* when it is $\approx$ mod M_n to a cycle on F_{n-1}. Then the difference between the cycle and its conjugate is ≈ 0, mod V_n.

17. Of no less importance than the preceding are the *skew-symmetric* cycles of V_n. We so designate those of type $C_\mu - \overline{C}_\mu$, where C_μ is on M_n. In place of them it would be possible to consider the complexes whose boundary is on F_{n-1} and their properties in regard to what might be called "quasi-homologies" or relations:

$$\sum t_i C^i + \text{a complex of } F_{n-1} \sim 0 \pmod{M_n}. \tag{17.1}$$

---

* Klein, *Elliptische Modulfunctionen*, vol. 2, p. 543.

The theory resulting therefrom would be largely a paraphrase of the one to be found here, with the mild advantage of being strictly confined to elements of M_n itself.

Since $C_\mu - \overline{C}_\mu$ is a cycle, the boundary of C_μ is on F_{n-1}. Furthermore we can remove from it any μ-cell on F_{n-1} without changing the cycle.

If ρ_μ is the μth connectivity index of V_n, any $\rho_\mu+1$ skew-symmetric μ-cycles are dependent, so that Klein's reasoning applies and we find a fundamental set $\Delta_\mu^1, \Delta_\mu^2, \cdots, \Delta_\mu^q$, for the type. It is reducible in similar fashion to the above as regards homologies between the intersections with the boundary:

$$\sum t_i \Delta_\mu^i \cdot F_{n-1} \sim 0 \pmod{F_{n-1}}, \tag{17.2}$$

with a similar conclusion: The set can be replaced by a new one, for which the same designation is preserved, with fundamental homologies for the type (17.2):

$$\tau_j \Delta_\mu^j \cdot F_{n-1} \sim 0 \pmod{F_{n-1}} \qquad (j = 1,2, \cdots, s \leqq q). \tag{17.3}$$

In short the first s cycles of the new set intersect the boundary of M_n in zero-divisors or bounding cycles of it, while the remaining $q-s$ intersect it in independent cycles of F_{n-1}. By Tr., No. 35, Theorem V, they are independent for V_n as well. There may exist, however, homologies between the first s. Reducing again as regards these we finally obtain a fundamental set consisting of the following:

(a) r_μ^3 cycles $\Delta_\mu^1, \Delta_\mu^2, \cdots, \Delta_\mu^{r_\mu^3}$, independent (mod V_n), but meeting F_{n-1} in cycles ≈ 0 (mod F_{n-1});

(b) $r_\mu^4 = q-s$ cycles $D_\mu^1, D_\mu^2, \cdots, D_\mu^{r_\mu^4}$, independent (mod V_n) and intersecting F_{n-1} in cycles independent (mod F_{n-1});

(c) a set of at most $q - r_\mu^3 - r_\mu^4$ zero-divisors of V_n.

Every skew-symmetric cycle is $\approx$ to a sum of Δ's and D's (these notations are henceforth reserved for cycles (a) and (b)). Also there can be no homology involving both D's and Δ's, as we see at once by reference to their intersections with F_{n-1}. Since there is none involving each type alone, they constitute $r_\mu^3 + r_\mu^4$ independent cycles of V_n, and therefore a fundamental set as regards $\approx$ and skew-symmetric cycles.

Let $\gamma_\mu \approx \Delta_\mu + D_\mu$ be a skew-symmetric cycle, Δ_μ and D_μ being sums of cycles (a) or (b). Let $\delta_{n-\mu}$ be any cycle of F_{n-1}. From §1 follows, with indices computed as to F_{n-1},

$$((\gamma_\mu \cdot F_{n-1}) \cdot \delta_{n-\mu}) = ((D_\mu \cdot F_{n-1}) \cdot \delta_{n-\mu}). \tag{17.4}$$

Hence, by §5 and the definition of the D's, in order that $\gamma_\mu \cdot F_{n-1} \approx 0 \pmod{F_{n-1}}$ it is necessary and sufficient that $\gamma_\mu \approx \Delta_\mu$ alone.

18. Let $\gamma_{\mu-1}$ be a cycle of F_{n-1} not ≈ 0 on that manifold, but bounding C_μ on M_n. Then $C_\mu - \overline{C}_\mu$ is a skew-symmetric cycle of type D not $\approx 0 \pmod{V_n}$, since its intersection γ_μ with F_{n-1} is not $\approx 0 \pmod{F_{n-1}}$. Hence to a set of t independent μ-cycles of F_{n-1} that bound on V_n correspond as many independent skew-symmetric cycles of type D, and conversely. Therefore r_μ^4 is the *number of distinct μ cycles F_{n-1} that bound on M_n.*

Another interesting property is the following: *Every Δ has a multiple which is the sum of a cycle on M_n and of a cycle on $\overline{M}_n$.* Let Δ_μ be the cycle, Δ_μ' a polyhedral approximation of maximum generality intersecting F_{n-1} in $\gamma_{\mu-1}$. By §17,

$$\gamma_{\mu-1} \approx 0 \pmod{F_{n-1}}. \tag{18.1}$$

Since Δ_μ' is of maximum generality it has no μ-cells on F_{n-1}, hence we may write

$$\Delta_\mu' = C_\mu - \overline{C}_\mu', \tag{18.2}$$

where the first complex is on M_n, the second on $\overline{M}_n$, and both have the common boundary $\gamma_{\mu-1}$. From (18.1) we infer that there exists on F_{n-1}

$$C_\mu'' \equiv t\gamma_{\mu-1}, \qquad t \neq 0, \tag{18.3}$$

therefore

$$t\Delta_\mu \sim t\Delta_\mu' = (tC_\mu - C_\mu'') - (t\overline{C}_\mu' - C_\mu''). \tag{18.4}$$

Each parenthesis at the right is a cycle, the first on M_n, the second on $\overline{M}_n$, which proves our assertion.

19. **Theorems.** I. *Every cycle of V_n is the sum of a skew-symmetric cycle and of one on M_n.*

Let γ_μ be the cycle. It may be assumed polyhedral and the sum of two complexes C_μ and $\overline{C}_\mu'$, the first on M_n, the second on $\overline{M}_n$. But

$$\gamma_\mu = (C_\mu + C_\mu') + (\overline{C}_\mu' - C_\mu'). \tag{19.1}$$

The second parenthesis is a skew-symmetric cycle, the first a complex on M_n, the difference of two cycles, hence also a cycle, and the theorem is therefore proved.

II. *A γ_μ of M_n not $\approx 0 \pmod{M_n}$ cannot be skew-symmetric.*

For let it be $\approx \delta_\mu$, skew-symmetric. Then

$$\delta_\mu - \gamma_\mu \approx 0 \approx \bar{\delta}_\mu - \bar{\gamma}_\mu \approx + \delta_\mu + \bar{\gamma}_\mu \pmod{V_n}, \tag{19.2}$$

Therefore

$$t(\gamma_\mu + \bar{\gamma}_\mu) \sim 0 \pmod{V_n}; \; t \neq 0. \tag{19.3}$$

First let γ_μ be symmetric, with $\theta\gamma_\mu$, $\theta \neq 0$, homologous (mod M_n) to γ_μ^0 on F_{n-1}. Then

$$\theta\gamma_\mu \sim \theta\bar{\gamma}_\mu \sim \gamma_\mu^0 \pmod{V_n}, \tag{19.4}$$

and therefore by (19.3)

$$2t\gamma_\mu^0 \sim 0 \pmod{V_n}. \tag{19.5}$$

There exists then on V_n a

$$C_{\mu+1} \equiv 2t\gamma_\mu^0. \tag{19.6}$$

Let $C'_{\mu+1}$ be the subcomplex of $C_{\mu+1}$ that includes all its cells on M_n (logical intersection of $C_{\mu+1}$ and M_n) and set $C_{\mu+1} - C'_{\mu+1} = \overline{C}''_{\mu+1}$, complex made up of all $(\mu+1)$-cells of $C_{\mu+1}$ interior to $\overline{M}_n$ plus their boundaries. Since the boundary of $\overline{C}''_{\mu+1}$ can only be on F_{n-1}, it coincides with that of $C''_{\mu+1}$. Hence

$$C'_{\mu+1} + C''_{\mu+1} \equiv 2t\gamma_\mu^0, \tag{19.7}$$

for $C_{\mu+1}$ has the same boundary as the complex at the left. But the latter is on M_n, hence

$$\theta\gamma_\mu \sim \gamma_\mu^0 \approx 0; \quad \gamma_\mu \approx 0 \pmod{M_n}, \tag{19.8}$$

contrary to assumption.

Assume now that γ_μ is not symmetric. It is reducible to a polyhedral cycle whose μ-cells are all interior to M_n. The first part of this double assertion is established as in Coll. Lect., pp. 95, 118, the second as in Tr. No. 24, (b). Furthermore a complete proof, independent of the present discussion, is given below (§23). The left member of (19.3), polyhedral and without μ-cells on F_{n-1}, will also bound a polyhedral $C_{\mu+1}$ (Coll. Lect., p. 120). Let again the sum of its cells on M_n be called $C'_{\mu+1}$. This last complex has for total boundary $t\gamma_\mu$ plus a cycle γ'_μ on F_{n-1}. Therefore

$$t\gamma_\mu \sim -\gamma'_\mu \pmod{M_n}, \tag{19.9}$$

and γ_μ is a symmetric cycle, — a new contradiction, and II is proved.

In the second part of the discussion we have established that if (19.3) holds then γ_μ is symmetric. The identical reasoning holds for the more general homology, in which γ_μ is still on M_n,

(19.10) $$t\gamma_\mu + \theta\bar{\gamma}_\mu \sim 0 \pmod{V_n}.$$

which shows that

III. *If γ_μ of M_n is dependent upon its conjugate γ_μ then it is a symmetric cycle.*

IV. *When γ_μ of M_n is not ≈ 0, mod M_n, then it is also not ≈ 0, mod V_n.*

This is a special case of II corresponding to $\delta_\mu = 0$.

20. From the preceding propositions follows at once the all important

THEOREM. *The cycles $\Gamma_\mu^{\alpha_1}$, $G_\mu^{\alpha_2}$, $\Delta_\mu^{\alpha_3}$, $D_\mu^{\alpha_4}$ $(\alpha_i = 1, 2, \cdots, r_\mu^i)$ constitute a fundamental set for V_n and the operation $\approx$. To obtain a fundamental set for $\sim$ it is only necessary to add zero-divisors of M_n and skew-symmetric zero-divisors of V_n.*

From their definition we know that the Γ's and G's are independent, and similarly for the Δ's and D's. From II follows that the four sets are independent in their totality. Then again from I and the fact that the Γ's and G's constitute a fundamental set as to $\approx$ and M_n, and similarly the Δ's and D's for the skew-symmetric cycles and V_n, the theorem follows in its completeness.

COROLLARY. *The μth connectivity index of V_n is $\rho_\mu = r_\mu^1 + \cdots + r_\mu^4$.*

21. Our present object is to show that with a suitable choice of associated sets for the dimensionalities μ and $n-\mu$ certain indices are, or can be made to be, zero, which will naturally lead to the canonical sets.

22. LEMMA. *Let E_n be a cell, Γ_μ a cycle on E_n, both polyhedral. Then there exists a polyhedral $C_{\mu+1}$ on E_n bounded by Γ_μ.*

Since Γ_μ is on E_n, it has no points on the boundary of the cell. Hence there exists a cell E_n' which together with its boundary lies on E_n, and also carries Γ_μ. Let σ be the least distance between points of the boundaries of the cells. Cover E_n with a polyhedral complex whose cells are all of diameter $<\sigma$, and remove all its n-cells with a boundary point on the boundary of E_n. The n-cells that are left together with their boundaries constitute a C_n on E_n and carrying E_n'. The cycle Γ_μ bounds on E_n', hence also on C_n, and on that complex it bounds a polyhedral $C_{\mu+1}$ (Coll. Lect., p. 120) which proves the lemma.

23. Let C_μ be a complex on M_n with its boundary on F_{n-1}, and let us follow step by step the approximation described in Tr., Part I, §4. We first apply the Alexander-Veblen process, and obtain $C_{\mu+1}$, polyhedral with an

associated congruence

$$C_{\mu+1} \equiv C_\mu - C'_\mu + C^0_\mu . \tag{23.1}$$

By reference to Coll. Lect., pp. 95, 118, we find that we may approximate each vertex of the boundary of C_μ by a vertex of the boundary of C_n, that is, by a point of F_{n-1}. Then C^0_μ and the boundary of C'_μ will be on F_{n-1}. Except for that, the rest of the work of approximation is directly applicable here. However, No. 23, Tr., does not apply to non-boundary cells of C'_μ on F_{n-1}, and necessitates a slight modification.

We take C_n such that C'_μ is now a subcomplex of it (Tr., No. 14, Lemma II), and as a first move, reduce its boundary on F_{n-1} as described in Tr., No. 23. We thus obtain a polyhedral cycle $\Gamma'_{\mu-1}$ of F_{n-1} whose $(\mu-i)$-cells are all on cells of no less than $n-i$ dimensions of C_n. Furthermore $\Gamma_{\mu-1}-\Gamma'_{\mu-1}$ bounds a polyhedral C''_μ, and both new cycles and complex are as near as we please to C'_μ, hence to C_μ. It follows that in (23.1), C'_μ and C^0_μ may be replaced by $C'_\mu+C''_\mu$ and $C^0_\mu+C''_\mu$, without altering the situation. Therefore we may start with a complex C'_μ whose boundary is already reduced as indicated. To extend then the reduction of Tr., No. 23, all we need to do is to replace C'_μ by a complex whose non-boundary cells are interior to M_n, since the reduction in question can be applied to these. The reader will verify with ease that the situation pertaining to (23.1) remains unaffected by any step to be taken presently. Furthermore, the reduction will be more thorough than in Tr., No. 23, in that the new elements introduced first here, then by the process of Tr., No. 23, will be throughout interior to M_n. Hence the complex as finally reduced will have all non-boundary cells interior to the manifold.

Let first E_μ be a simplicial cell of C'_μ on F_{n-1} and on the boundary of the cell E_n of C_n. Draw rectilinear segments from a fixed point of E_n to all points of E_μ. The resulting simplicial cell has for boundary a Γ_μ whose μ-cells other than E_μ (which is one of them) are on E_n. There is a $t\neq 0$ such that $C'_\mu - t\Gamma_\mu \sim C'_\mu$ is a complex which no longer includes E_μ. The cell E_μ has then been replaced by cells on E_n, or interior cells of M_n. Thus we can reduce C'_μ to a similar complex whose μ-cells are interior to M_n. Assume then that all cells of more than h dimensions, $h<\mu$, of C'_μ are interior cells. I say that the reduction can be extended to the h-cells as well.

Let E_h of C'_μ be on F_{n-1}. Introduce new interior vertices on C_n so chosen that for the new complex C'_n the star of cells of center E_n carries no boundary cells of C'_μ on its own. Since the star attached to C_n carries no other boundary cells of C'_μ than those on E_h or its boundary, the construction offers no

difficulty. The related C_{n-h-1} of Tr., No. 3, which is merely the boundary of the star, will then intersect C'_μ in a $(\mu-h-1)$-cycle on E_{n-h-1}. By our lemma this cycle bounds a polyhedral $C_{\mu-h}$ on E_{n-h-1}, having then no point on F_{n-1}. Let $E_h = A_0 A_1 \cdots A_h$, and let also $A_{h+1} \cdots A_i$ be an arbitrary cell of $C_{\mu-h}$; then denote by $\overline{C}_{\mu+1}$ the complex sum of the cells $A_0 A_1 \cdots A_i$. It has in common with C'_μ all cells incident with E_h, and except for these cells and their boundary points it is entirely interior to M_n. Let us sense its boundary Γ'_μ so that the indicatrix of any cell is obtained by naming first the vertices of E_h, next those of the related cell on C_{n-h-1} so that both sets be vertices of an indicatrix for their cell. Then C'_μ and Γ'_μ will have the same cells incident with E_h each counted with the same multiplicity for both complexes. Hence $C'_\mu - \Gamma'_\mu \sim C'_\mu$ will have lost E_h without acquiring new cells on F_{n-1}. This shows that the reduction can also be extended to h cells of C'_μ, hence to its boundary cells.

Combining the whole discussion we have the important

THEOREM. *Let C_n be an assigned defining complex of M_n and C_μ a complex on the manifold, whose boundary is on F_{n-1}. Then there is a corresponding congruence* (23.1) *with* (a) *C'_μ polyhedral, as near as we please to C_μ, with its boundary on F_{n-1} and as near as we please to that of C_μ, also with its $\mu-i$ cells on cells of no less than $n-i$ dimensions of C_n, the non-boundary cells being interior to M_n*; (b) *C^0_μ on F_{n-1} and as near as desired to the boundary of C_μ.*

COROLLARY I. *Every cycle of M_n is homologous to an interior cycle behaving in accordance with the theorem.*

COROLLARY II. *Every skew-symmetric cycle is homologous to one of form $C'_\mu - \overline{C}'_\mu$, where C'_μ behaves in accordance with the theorem.*

For let $\gamma_\mu = C_\mu - \overline{C}_\mu$. Reduce C_μ as above with the congruence (23.1). Then also

$$\overline{C}_{\mu+1} \equiv \overline{C}_\mu - \overline{C}'_\mu + C^0_\mu \; ; \tag{23.2}$$

therefore

$$\gamma_\mu - (C'_\mu - \overline{C}'_\mu) \sim 0, \tag{23.3}$$

as was to be proved.

Unless otherwise stated we shall always assume the cycles of M_n and the skew-symmetric cycles reduced as far as allowed by theorem and corollaries. Practically everywhere in the sequel, a cycle of one of these two types may be replaced by one which is $\sim$, mod M_n and V_n respectively, and then

if it is not already reduced, we shall be at liberty to reduce it without further discussion.

24. **Theorems on indices.** I. $(\Gamma_\mu^i \cdot G_{n-\mu}^j) = (G_\mu^i \cdot G_{n-\mu}^j) = 0$.

There exists $G_{n-\mu}$ of F_{n-1}, $\sim t\, G_{n-\mu}^j$, mod M_n, $t \neq 0$. If we replace $G_{n-\mu}^j$ by $G_{n-\mu}$ we merely multiply the indices by t. Hence we need only show that

$$(24.1) \qquad (\Gamma_\mu^i \cdot G_{n-\mu}) = (G_\mu^i \cdot G_{n-\mu}^j) = 0,$$

which is obvious since Γ_μ^i and G_μ^i may be chosen interior to M_n and then they will not meet $G_{n-\mu}$.

Explicitly, and since $(\Gamma \cdot G) = \pm (G \cdot \Gamma)$, *the indices* $(\Gamma \cdot G)$, $(G \cdot \Gamma)$, $(G \cdot G)$ *are all zero.*

II. *The index of two skew-symmetric cycles is zero.*

Let $\delta_\mu = C_\mu - \overline{C}_\mu$, $\delta_{n-\mu} = C_{n-\mu} - \overline{C}_{n-\mu}$ be the two cycles in the reduced form in position of maximum generality. Then for evident reasons of symmetry the two indices $(C_\mu \cdot C_{n-\mu})$ and $(\overline{C}_\mu \cdot \overline{C}_{n-\mu})$ taken with M_n and $\overline{M}_n$ as the carrying manifolds are equal. Hence, taken with $V_n = M_n - \overline{M}_n$ as the carrying manifold, they are opposite. But C_μ does not meet $\overline{C}_{n-\mu}$ and $C_{n-\mu}$ does not meet $\overline{C}_\mu$. Hence

$$(24.2) \qquad \begin{aligned} (\delta_\mu \cdot \delta_{n-\mu}) &= (C_\mu - \overline{C}_\mu)(C_{n-\mu} - \overline{C}_{n-\mu}) \\ &= (C_\mu \cdot C_{n-\mu}) - (\overline{C}_\mu \cdot \overline{C}_{n-\mu}) = 0. \end{aligned}$$

III. $(G_\mu^i \cdot \Delta_n^j{}_{-\mu}) = 0$.

In the proof G may be replaced by $t\,G$, $t \neq 0$. Since there is a $t\,G$ on F_{n-1}, we may assume that G itself is on the boundary. Then (§ 18) there is an $s \neq 0$ such that

$$(24.3) \qquad s\Delta \sim \Delta' + \overline{\Delta}'',$$

where Δ' is a cycle on M_n and $\overline{\Delta}'$ a cycle on $\overline{M}_n$. By I and since Δ' and Δ'' depend upon the G's and Γ's,

$$(24.4) \qquad (G \cdot \Delta') = (G \cdot \Delta'') = -(G \cdot \overline{\Delta}'') = 0\,;$$

therefore

$$(24.5) \qquad (G \cdot \Delta) = 0 = (\Delta \cdot G).$$

Conclusion. *All indices* $(\Gamma \cdot G)$, $(\Delta \cdot G)$, $(\Delta \cdot \Delta)$, $(\Delta \cdot D)$, *and those obtained by permuting the cycles, are zero.*

25. The matrix L_μ will then have the form

$$L_\mu = \begin{array}{c|cccc|c} & \Gamma_{n-\mu} & \Delta_{n-\mu} & G_{n-\mu} & D_{n-\mu} & \\ \hline \| & A & B & 0 & C & \| \Gamma_\mu \\ \| & H & 0 & 0 & 0 & \| \Delta_\mu \\ \| & 0 & 0 & 0 & F & \| G_\mu \\ \| & M & 0 & N & 0 & \| D_\mu \end{array}$$

where the terms are all matrices; for example, $H = ||(\Gamma_\mu^i \cdot \Delta_n{}^j{}_{-\mu})||$, $||(\Delta_\mu^i \cdot \Delta_n{}^j{}_{-\mu})|| = 0$, etc. We shall now study L_μ and in particular show that by proper choice of cycles in each group it can be reduced to a much simpler form, with only one indeterminate matrix, namely A. As an incidental result we shall obtain very interesting duality theorems regarding the integers r_μ, theorems quite similar to Poincaré's relation for the connectivity numbers of manifolds without boundary.

26. I. *F and N are square.* The permissible operations of adding a multiple of a cycle of a fundamental set to another or permuting two of them, applied to the groups G_μ, $D_{n-\mu}$ will amount to the noted elementary transformations as applied to F. Hence F may be reduced to the well known form

(26.1)
$$\begin{array}{c} \\ \\ \\ \\ p \\ \\ \\ \end{array} \overbrace{\begin{Vmatrix} e_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & e_2 & \cdots & & & & \\ \cdots & & & & & & \\ 0 & \cdots & & e_s & 0 & \cdots & 0 \\ 0 & \cdots & & 0 & 0 & \cdots & 0 \\ \cdots & & & & & & \\ 0 & \cdots & & & & & 0 \end{Vmatrix}}^{q}$$

where the e's are the invariant factors of F and there are p rows and q columns of zeros. Assume this done and continue to call the reduced matrix F. We must show that $p = q = 0$. Evidently, $p = 0$, for otherwise L_μ, as reduced, would have a whole row of zeros, whereas its determinant is ± 1 (§ 6). Then if $q \neq 0$ there is a $D_{n-\mu}$ such that $(G_\mu \cdot D_{n-\mu}) = 0$ for every symmetrical cycle G_μ and in particular for every one on F_{n-1}. Hence (§§5, 17), $D_{n-\mu}$ is dependent upon the cycles $\Delta_{n-\mu}$, which is untrue. Therefore $q = 0$, F is square and so similarly is N. Their determinants are integers and factors of $|L_\mu| = \pm 1$, hence they are both ± 1.

In the canonical form the e's for both matrices will be $+1$. When $\mu \neq \frac{1}{2}n$, they may both be separately reduced to that form. Denoting generically by I_k the unit matrix of order k, we have, when the reduction is carried

out, $F=I_{r_\mu^2}$, $N=I_{r_\mu^4}$. When $\mu=\frac{1}{2}n$, we have at once, by Tr., No. 8, $N=(-1)^{n/2}F'$. The reduction of F brings about that of N. For $\frac{1}{2}n$ even, the final form is as above; for $\frac{1}{2}n$ odd, it is $F=I_{r_\mu^2}=-N$.

As an important result proved incidentally we have $r_\mu^2=r_{n-\mu}^4$ or

II. (FIRST DUALITY THEOREM.) *The number of distinct symmetric μ cycles is equal to the number of independent $n-\mu$ cycles of F_{n-1} that bound on M_n.*

Since F is a unit matrix and since a $\Gamma-G$ is also a Γ, we can subtract from every Γ_μ of the fundamental set a G_μ cycle so chosen as to reduce the corresponding row of C to zero. This means that we can so select the set of Γ_μ's as to have $C=0$. If $\mu\neq\frac{1}{2}n$, we may operate similarly on the $\Gamma_{n-\mu}$'s and reduce M to zero, while when $\mu=\frac{1}{2}n$, we shall have $M=\pm C=0$. Hence

III. *The fundamental sets can be so selected as to give L_μ one of the two forms*

$$\left\|\begin{matrix} A & B & 0 & 0 \\ H & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{matrix}\right\|, \qquad \left\|\begin{matrix} A & B & 0 & 0 \\ H & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{matrix}\right\|,$$

according as μ is not or is $\frac{1}{2}n$ and odd. To simplify we have merely indicated the unit matrices by 1.

IV. *A is a square matrix.* It may be reduced to the type (26.1) by the two operations of the beginning of this section applied to Γ's alone. As reduced to that type we shall show again that $p=q=0$. For evident reasons of symmetry it is sufficient to show that $q=0$. For every Γ_μ we have $(\Gamma_\mu\cdot\Gamma_{n-\mu}^{s+1})=0$. Then if $\delta_{n-\mu}=\Gamma_{n-\mu}^{s+1}-\overline{\Gamma}_{n-\mu}^{s+1}$, we have, for every G_μ,

$$(26.2) \qquad (G_\mu\cdot\delta_{n-\mu}) = (G_\mu\cdot\Gamma_{n-\mu}^{s+1}) - (G_\mu\cdot\overline{\Gamma}_{n-\mu}^{s+1}) = 0,$$

for the last two indices are equal in absolute value and the first is zero (§ 24, Theorem I). Also, as we may assume Γ_μ interior to M_n and $\overline{\Gamma}_{n-\mu}^{s+1}$ interior to $\overline{M}_n$, and hence that the two are without common points,

$$(26.3) \qquad (\Gamma_\mu\cdot\delta_{n-\mu}) = -(\Gamma_\mu\cdot\overline{\Gamma}_{n-\mu}^{s+1}) = 0.$$

And finally, since $\delta_{n-\mu}$ is skew-symmetric,

$$(26.4) \qquad (\Delta_\mu\cdot\delta_{n-\mu}) = (D_\mu\cdot\delta_{n-\mu}) = 0.$$

In short, $(\gamma_\mu\cdot\delta_{n-\mu})=0$ for every μ cycle of the fundamental set, therefore for every μ cycle of V_n. Hence $\delta_{n-\mu}\approx 0$. There exists, then, a $t\neq 0$ such that $t\Gamma_{n-\mu}^{s+1}\sim t\overline{\Gamma}_{n-\mu}^{s+1}$. Hence (§ 19, Theorem III), $\Gamma_{n-\mu}^{s+1}$ is a symmetric cycle,

contrary to assumptions. There exists then no $\Gamma_{n-\mu}$ whose upper index $>s$, which means that $q=0$ and proves IV.

Since A is square, $r_\mu^1=r_{n-\mu}^1$ and therefore

V. (SECOND DUALITY THEOREM.) *The number of cycles of M_n of which no combination is a cycle of F_{n-1} is the same for the dimensions μ and $n-\mu$.*

VI. *B and H are square matrices and their determinants are ± 1.* We reduce again say B to the form (26.1) and show that $p=q=0$. If $q\neq 0$ there exists a $\Delta_{n-\mu}$ such that $(\gamma_\mu\cdot\Delta_{n-\mu})=0$ whatever γ_μ, hence $\Delta_{n-\mu}$ is a zero-divisor contrary to assumptions, and $q=0$. Assume now $p\neq 0$. Then the number of distinct $\Delta_{n-\mu}$'s is $<r_\mu^1=r_{n-\mu}^1$, the order of A. Consider, however, the cycles $\Gamma_{n-\mu}^i-\overline{\Gamma}_{n-\mu}^i$. Since the Γ's may be taken interior to M_n, these skew-symmetric cycles do not intersect F_{n-1} and therefore are dependent upon the Δ's. On the other hand they are independent, or there would exist a $\Gamma_{n-\mu}\approx\overline{\Gamma}_{n-\mu}$, and hence symmetric (§ 19, Theorem III) in contradiction to the assumptions on the Γ type. Hence there are at least r_μ^1 distinct $\Delta_{n-\mu}$'s and $p=0$. Therefore B is a square matrix and so is H. Here again their determinants are integers and factors of $|L_\mu|=\pm 1$, therefore also $=\pm 1$.

From VI it follows that $r_\mu^1=r_{n-\mu}^1=r_\mu^3=r_{n-\mu}^3$. Hence

VII. (THIRD DUALITY THEOREM.) *The number of distinct skew-symmetric cycles that intersect F_{n-1} in zero-divisors or bounding cycles is the same for the dimensions μ and $n-\mu$ and equal to the number of distinct cycles of M_n of μ or $n-\mu$ dimensions that are independent of the cycles of F_{n-1}.*

27. It follows from the above that for $\mu\neq\frac{1}{2}n$ we can reduce both B and H to $I_{r_\mu^1}$. For $\mu=\frac{1}{2}n$, $H=(-1)^n B$ and the reduction of B will bring about that of H.

For the computation to follow it is advisable to select fundamental sets thus: when $\mu<\frac{1}{2}n$ the four groups of cycles are taken in the order Γ, Δ, D, G; when $\mu\geqq\frac{1}{2}n$ in the order Γ, Δ, G, D. Then

$$(27.1)\qquad \mu\neq\frac{1}{2}n,\quad L_\mu=\left\|\begin{array}{cc|c} A & 1 & \\ 1 & 0 & 0 \\ \hline 0 & & 1\ 0 \\ & & 0\ 1 \end{array}\right\|;$$

$$\mu=\frac{1}{2}n,\quad L_\mu=\left\|\begin{array}{cc|c} A & 1 & \\ (-1)^{n/2}\cdot 1 & 0 & 0 \\ \hline 0 & & 0\ 1 \\ & & (-1)^{n/2}\cdot 1\ 0 \end{array}\right\|$$

where the 1 stands in each case for a unit matrix whose order is not essential (see Remark below). What matters chiefly is that for the dimensions μ and $n-\mu$ those in the same places are of the same orders. In the upper left corner they are both of order r_μ^1, in the right corner of orders r_μ^2 and r_μ^4. Concerning A we have no available information, but fortunately it disappears entirely from our formulas and therefore need not concern us further.

Remark. Consider for a moment two matrices written in the form

$$\alpha = \| \alpha_{ij} \|, \qquad \beta = \| \beta_{ij} \|, \tag{27.2}$$

with elements α_{ij}, β_{ij} themselves matrices. The ordinary multiplication rule

$$\alpha\beta = \| \sum \alpha_{ik}\beta_{kj} \| \tag{27.3}$$

is directly applicable (taking care not to interchange factors in $\alpha_{ik}\,\beta_{kj}$) provided that (a) the number of columns in α is the same as the number of rows in β; (b) the sequence of the number of columns for the elements in a row of α is the same for all rows and also the same as for β'. These two conditions are fulfilled if α and β are square with their diagonal elements α_{ii}, β_{ii} also square and of equal order for the same i. In that case not only $\alpha\beta$ but also $\beta\alpha$ may be obtained by the usual rule. This is the precise situation that we shall face throughout, where we shall find products of matrices all of the same structure as L_μ.

Let us recall incidentally that, for example, $\alpha' = \| \alpha_{ji}' \|$, that is, the transposed of α is obtained by interchanging rows and columns and replacing each individual term by its transposed. All this goes back to the rule for matrix multiplication.

III. Continuous transformations of manifolds with a boundary

28. Just as in the no-boundary case the definition of a continuous transformation is best given by reference to $M_n \times M_n'$, where M_n' is a copy of M_n. However, before discussing the transformations we shall show that the product is also a manifold. The M_n' image of any M_n configuration will be denoted throughout by the same letter accented.

Let then E_h be any cell of C_n, defining complex of M_n, s_{n-h} the corresponding system such as appears in § 12 in connection with conditions (a′), (b′). We have

$$s_{n-h} = E_{n-h} + tE_{n-h-1}, \tag{28.1}$$

where $t=0$ or 1 according as E_h is or is not an interior cell of C_n. To any cell E_k of C_n' will similarly correspond

$$s_{n-k} = E_{n-k} + t'E_{n-k-1}, \tag{28.2}$$

where all terms have an obvious meaning. But referring to the definition of the s's we find that $s_{n-h}\times s_{n-k}$ corresponds in a similar manner to $E_h\times E_k$ as a cell of $C_n\times C_n'$. Now

$$\begin{aligned} s_{n-h}\times s_{n-k} = {} & E_{n-h}\times E_{n-k} + tE_{n-h-1}\times E_{n-k} \\ & + t'E_{n-h}\times E_{n-k-1} + tt'E_{n-h-1}\times E_{n-k-1}. \end{aligned} \tag{28.3}$$

Each term represents a cell whose dimensions are the sum of those of the factors. Now $E_h\times E_k$ is a boundary cell only when one of the t's is not zero. When both are zero, (a') is manifestly satisfied, and when, say, $t=1$, $t'=0$, (b') is satisfied, as they should be. Let, then, $t=t'=1$. We must show that $E_{n-h-1}\times E_{n-k}+E_{n-h}\times E_{n-k-1}+E_{n-h-1}\times E_{n-k-1}$ is homeomorphic to a $2(n-h-k-1)$-cell. It will be remembered that, by condition (b'), E_{n-k} is homeomorphic to the interior of a hemisphere in S_{n-k} with E_{n-k-1} as the flat base of the hemisphere. It follows that the first term is of the same type for an $S_{2n-h-k-1}$ and similarly for the second with the third as the common flat base. The sum is then homeomorphic to the interior of a sphere in the same space, that is to a cell, which completes the proof.

Remarks. I. The same proof holds for a product $M_p\times M_q$, $p\neq q$.

II. Since the factors are orientable this is also true for the product (Tr., No. 49).

29. We now define a continuous transformation T of M_n as in Tr., No. 56, by the condition that the set $\{A\times B\}=K_n$ be a C_n of $M_n\times M_n'$ with its boundary on that of the product. Let T' be another transformation, K_n' its complex. The problem is again to determine $(K_n\cdot K_n')$ (number of signed coincidences) in terms of the transformations induced by T, T' on the cycles, or of similar data (i. e., information naturally at hand when T, T' are known).

No result of any generality is to be expected unless K_n and K_n' are so restricted that the boundaries of suitably defined approximations do not intersect. Indeed, unless this is so, $(K_n\cdot K_n')$ ceases to be an invariant of classes of transformations. We assume of course throughout that M_n is connected, but F_{n-1} need not be so. Let $F_{n-1}^1, \cdots, F_{n-1}^p$ be its connected parts. By Tr., (49.2), the boundary of $M_n\times M_n'$ is the sum of the products $F_{n-1}^i\times M_n'$, $(-1)^n M_n\times F'^i_{n-1}$, each of which is a manifold (§ 28). The

boundary of each K consists of subcomplexes distributed among some or all of these 2 p manifolds. The least that we can exact is that none carry boundary points of both K's in its interior. However, this is still beyond the reach of the method to be used here, at least in this general form. *We shall therefore restrict our discussion to pairs of transformations such that the boundary of one K is on $F_{n-1}\times M_n'$, while that of the other is on $M_n\times F_{n-1}$, one of the boundaries being actually interior to the carrying manifold.* This will greatly simplify matters, sufficiently indeed to compensate amply for whatever may be lost in generality. In § 32 we shall give a topological interpretation of these types of transformations by means of certain approximating transformations.

30. We shall reduce the coincidence and fixed points problems for M_n to similar problems for certain associated transformations of V_n that we now define.

Case I: *T is of the first type*, that is with the boundary Γ_{n-1} of K_n on $F_{n-1}\times M_n'$. According to § 23, Corollary I, it is homologous thereon to Γ_{n-1}' interior to $F_{n-1}\times M_n'$ and satisfying in all respects the conditions there stated. If beyond this point we apply the same reductions as before (loc. cit.) to K_n itself, we shall reduce it to a complex H_n bounded by Γ_{n-1}' with its non-boundary cells all interior to $M_n\times M_n'$ and behaving in every respect in accordance with the theorem of § 23. If C_μ of M_n has its boundary on F_{n-1} we first reduce it as in § 23; then its transform TC_μ is determined as in Tr., No. 58, with C_μ in place of γ_μ and H_n in place of Γ_n (loc cit.). As a special case C_μ may be a γ_μ; then it is first reduced to the interior of M_n and $T\gamma_\mu$ is then determined in the same way.

Since $V_n\times V_n'=M_n\times V_n'-\overline{M}_n\times V_n'$, it may be derived from $M_n\times V_n'$ as V_n from M_n. The two parts of the manifold are now matched along their common boundary $F_{n-1}\times V_n'$, and their points associated in conjugate pairs $A\times B$ and $\overline{A}\times B$. To K_n and H_n there correspond in this fashion associated skew-symmetric cycles $K_n-\overline{K}_n\sim H_n-\overline{H}_n=\Gamma_n$, and the latter will serve to define the transformation T_1 of V_n associated with T.

Since the B points describe on V_n' the images of the transforms of the loci of the A points, we have, by Tr., No. 58,

$$(30.1)\qquad T_1\overline{C}_\mu = T_1C_\mu = TC_\mu\,; \qquad T_1(C_\mu-\overline{C}_\mu)=0.$$

Furthermore, $T_1\gamma_\mu=T\gamma_\mu$ since K_n' alone comes into play in determining the two transforms. Hence *T_1 transforms cycles of M_n into cycles of M_n and skew-symmetric cycles into bounding cycles.*

Let the transformation matrix for the cycles Γ_μ, G_μ of the fundamental sets be

$$P_\mu = \begin{array}{c} \begin{array}{cc} \Gamma_\mu & G_\mu \end{array} \\ \left\| \begin{array}{cc} P_{11}^\mu & P_{12}^\mu \\ P_{21}^\mu & P_{22}^\mu \end{array} \right\| \begin{array}{l} \Gamma_\mu \\ G_\mu \end{array} \end{array}.$$

It is the transformation matrix for the cycles of M_n and T, the transformed cycles being given as $\approx$ to a combination of the initial cycles. The similar matrix for V_n is

$$\mu < \frac{1}{2}n,\ \alpha_\mu = \begin{array}{c} \begin{array}{cccc} \Gamma & \Delta & D & G \end{array} \\ \left\| \begin{array}{cccc} P_{11}^\mu & 0 & 0 & P_{12}^\mu \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ P_{21}^\mu & 0 & 0 & P_{22}^\mu \end{array} \right\| \begin{array}{l} \Gamma \\ \Delta \\ D \\ G \end{array} \end{array}; \quad \mu \geqq \frac{1}{2}n,\ \alpha_\mu = \begin{array}{c} \begin{array}{cccc} \Gamma & \Delta & G & D \end{array} \\ \left\| \begin{array}{cccc} P_{11}^\mu & 0 & P_{12}^\mu & 0 \\ 0 & 0 & 0 & 0 \\ P_{21}^\mu & 0 & P_{22}^\mu & 0 \\ 0 & 0 & 0 & 0 \end{array} \right\| \begin{array}{l} \Gamma \\ \Delta \\ G \\ D \end{array} \end{array}.$$

31. Case II: *T is of the second type*, or with its boundary on $M_n \times F'_{n-1}$. In this case H_n will have its boundary interior to $M_n \times F'_{n-1}$. Then since $V_n \times V'_n = V_n \times M'_n - V_n \times \overline{M}'_n$, the left side is to be considered as obtained by matching the two manifolds at the right whose common boundary is $V_n \times F'_{n-1}$, and $A \times B$, $A \times \overline{B}$ constitute the conjugate pairs. Again $K_n - \overline{K}_n \sim H_n - \overline{H}_n = \Gamma_n$, cycle which serves to define T_1.

We find now that $T_1\, C_\mu$ consists of $C'_\mu = TC_\mu$ and $-\overline{C}'_\mu$, the second complex having a minus sign because its orientation is determined by means of $-\overline{K}'_n$. Also no cell on $\overline{M}_n$ has any transform. It follows that *every cycle of V_n is transformed into a skew-symmetric cycle by T_1*. More explicitly, let the cycle first polyhedrally approximated be $\gamma_\mu = C'_\mu + \overline{C}''_\mu$, where the first complex is on M_n, the second on $\overline{M}_n$. Then $T_1\, \gamma_\mu = TC'_\mu - (\overline{TC}'_\mu)$.

The transformation matrix is given below:

$$\mu < \frac{1}{2}n,\ \alpha_\mu = \begin{array}{c} \begin{array}{cccc} \Gamma & \Delta & D & G \end{array} \\ \left\| \begin{array}{cccc} 0 & R_{11}^\mu & R_{12}^\mu & 0 \\ 0 & Q_{11}^\mu & Q_{12}^\mu & 0 \\ 0 & Q_{21}^\mu & Q_{22}^\mu & 0 \\ 0 & R_{21}^\mu & R_{22}^\mu & 0 \end{array} \right\| \begin{array}{l} \Gamma \\ \Delta \\ D \\ G \end{array} \end{array};$$

$$\mu \geqq \frac{1}{2}n,\ \alpha_\mu = \begin{array}{c} \begin{array}{cccc} \Gamma & \Delta & G & D \end{array} \\ \left\| \begin{array}{cccc} 0 & R^\mu_{11} & 0 & R^\mu_{12} \\ 0 & Q^\mu_{11} & 0 & Q^\mu_{12} \\ 0 & R^\mu_{21} & 0 & R^\mu_{22} \\ 0 & Q^\mu_{21} & 0 & Q^\mu_{22} \end{array} \right\| \end{array} \begin{array}{c} \Gamma \\ \Delta \\ G \\ D \end{array}.$$

The matrix $Q = ||Q^\mu_{ij}||$ is the transformation matrix for skew-symmetric cycles, and alone will appear in the final formulas.

32. The complex H_n defines for both types an approximation $\overline{T}$ to the given T such that (a) when T is of the first type, $\overline{T}M_n$ is wholly interior to M_n; (b) when T is of the second type, $\overline{T}F_{n-1}$ does not exist, i.e., the boundary belongs to the set of points that have no $\overline{T}$ transform. This may be considered as a geometric characterisation of the two types.

33. Let us now assume that T is of type I, T' of type II and represent the cycle and complexes attached to T' by the same letters as for T with primes. We are explicitly assuming (§ 29) that K and K' intersect only in interior points of $M_n \times M_n'$. Hence $(K_n \cdot K_n')$ is perfectly determined and by definition equal to $(H_n \cdot H_n')$ (Tr., No. 35). Therefore, at once,

$$(33.1) \qquad (\Gamma_n \cdot \Gamma_n') = ((H_n - \overline{H}_n) \cdot (H_n' - \overline{H}_n')) = (H_n \cdot H_n') = (K_n \cdot K_n'),$$

for when a term of a pair H, H' is barred the two complexes do not meet. Hence *the coincidence problem for T, T' is reduced to the same problem for T_1, T_1'.*

IV. Coincidence and fixed points formulas

34. We have just shown that these formulas are the same for M_n as for the associated transformations of V_n. We apply then (10.4), taking for T_1 the transformation corresponding to the α's, for T_1' that corresponding to the β's, and we have

$$(34.1) \qquad (K_n \cdot K_n') = (\Gamma_n \cdot \Gamma_n') = \sum (-1)^\mu \operatorname{trace} L_\mu \beta'_{n-\mu} L_\mu^{-1} \alpha_\mu.$$

We now assume the fundamental sets in the canonical form of § 27 and carry out the computation. The matrices in the product have the same structure, with four square submatrices in the principal diagonal in the same order. L_μ^{-1} for $\mu \neq \frac{1}{2}n$ odd, is like L_μ except that

$$\left\| \begin{array}{cc} A & 1 \\ 1 & 0 \end{array} \right\| \text{ is replaced by } \left\| \begin{array}{cc} 0 & 1 \\ 1 & -A \end{array} \right\|.$$

For $\mu=\frac{1}{2}n$ odd,

$$L_{n/2}^{-1}=\left\|\begin{array}{cc|cc} 0 & -1 & & \\ 1 & A & 0 & \\ \hline & & 0 & -1 \\ & 0 & 1 & 0 \end{array}\right\|. \tag{34.2}$$

By way of illustration, we examine the computation for $\mu<\frac{1}{2}n$. Then, dropping the indices μ and $n-\mu$ for the present,

$$L\beta'=\left\|\begin{array}{cc|c} A & 1 & \\ 1 & 0 & 0 \\ \hline & 0 & 1 \end{array}\right\| \cdot \left\|\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ R'_{11} & Q'_{11} & R'_{21} & Q'_{21} \\ \hline 0 & 0 & 0 & 0 \\ R'_{12} & Q'_{12} & R'_{22} & Q'_{22} \end{array}\right\|$$

$$=\left\|\begin{array}{cc|cc} R'_{11} & Q'_{11} & R'_{21} & Q'_{21} \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ R'_{12} & Q'_{12} & R'_{22} & Q'_{22} \end{array}\right\|; \tag{34.3}$$

$$L^{-1}\alpha=\left\|\begin{array}{cc|c} 0 & 1 & \\ 1 & -A & 0 \\ \hline & 0 & 1 \end{array}\right\| \cdot \left\|\begin{array}{cc|cc} P_{11} & 0 & 0 & P_{12} \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ P_{21} & 0 & 0 & P_{22} \end{array}\right\| \tag{34.4}$$

$$=\left\|\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ P_{11} & 0 & 0 & P_{12} \\ \hline 0 & 0 & 0 & 0 \\ P_{21} & 0 & 0 & P_{22} \end{array}\right\|.$$

Hence

$$L\beta'L^{-1}\alpha=\left\|\begin{array}{ccccccc} Q'_{11}P_{11}+Q'_{21}P_{21} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 0 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & Q'_{12}P_{12}+Q'_{22}P_{22} \end{array}\right\| \tag{34.5}$$

where the terms not in the main diagonal are omitted and need not be computed, as unnecessary for the trace. It follows that

$$\text{trace } L_\mu\beta'_{n-\mu}L_\mu^{-1}\alpha_\mu=\text{trace } \sum Q'^{n-\mu}_{ij}P^\mu_{ij}=\text{trace } Q'_{n-\mu}P_\mu, \tag{34.6}$$

therefore

$$(K_n \cdot K_n') = \sum(-1)^\mu \operatorname{trace} Q'_{n-\mu} P_\mu \tag{34.7}$$

is the desired coincidence formula. Its similarity to (10.4) for the non-boundary case is very striking.

35. The preceding formula corresponds to highly specialized fundamental sets. Let us consider the more general case where they are merely composed of the same four types Γ, Δ, G, D, in the same order as before for each μ, but no further specialized than is demanded by the condition that all indices $(\Gamma \cdot D)$, $(D \cdot \Gamma)$ be zero. This amounts to requiring that $C = M = 0$, where the two matrices are as in § 25. The passage from a canonical set to the more general type is by means of transformations operating separately on each of the four types. Then

$$L = \left\| \begin{array}{cc|cc} A & B & 0 & \\ H & 0 & & \\ \hline & & F & 0 \\ 0 & & 0 & N \end{array} \right\| \qquad (\mu < \tfrac{1}{2}n),$$

$$= \left\| \begin{array}{cc|cc} A & B & 0 & \\ H & 0 & & \\ \hline & & N & 0 \\ 0 & & 0 & F \end{array} \right\| \qquad (\mu > \tfrac{1}{2}n), \tag{35.1}$$

$$= \left\| \begin{array}{cc|cc} A & & B & \\ (-1)^{n/2}B & 0 & & 0 \\ \hline & 0 & & \\ 0 & & F & \\ (-1)^{n/2}F & & 0 & \end{array} \right\| \qquad (\mu = \tfrac{1}{2}n),$$

where all terms have the subscript μ omitted for the sake of simplicity. It turns out that the only terms needed are

$$B_\mu = \| (\Gamma_\mu^i \cdot \Delta_{n-\mu}^j) \|, \qquad F_\mu = \| (D_\mu^i \cdot G_{n-\mu}^j) \|. \tag{35.2}$$

Then, either by computing directly as before or else from (34.7), we derive the equivalent invariant form:

$$\begin{aligned} (K_n \cdot K_n') = \sum(-1)^\mu \operatorname{trace} (&B_\mu Q'^{n-\mu}_{11} B_\mu^{-1} P_{11} \\ &+ B Q'_{21} F^{-1} P_{21} + F Q'_{12} B^{-1} P_{12} + F Q'_{22} F^{-1} P_{22}), \end{aligned} \tag{35.3}$$

where the dimensionality indices, the same for all four terms at the right, are written only for the first. This is the desired generalization of (34.7).

36. We pass now to the derivation of the *fixed point formulas*. We assume the fundamental sets general in the sense of § 35.

I. *T is of the first type.* Since the identity is of each of the two types, we now assign it to the second type, approximate its K, say K_n^0, by H_n^0 with its boundary on $M_n \times F'_{n-1}$, and interpret the problem as the determination of $(K_n \cdot K_n^0)$. It is in fact exactly that when T actually reduces F_{n-1} to the interior of M_n.

We shall so sense K_n^0 that the corresponding integer θ of § 8, here the same at every point of K_n^0, is $+1$. Then it is immediately seen that the corresponding $Q_\mu = I$.

Whence

$$Q_{12}^{n-\mu} = Q_{21}^{n-\mu} = 0, \quad Q_{11}^{n-\mu} = Q_{22}^{n-\mu} = 1. \tag{36.1}$$

Therefore

$$(K_n \cdot K_n^0) = \sum(-1)^\mu \text{ trace } P_\mu. \tag{36.2}$$

II. *T is of the second type.* The discussion is now the same, with the identity ascribed to the first type. H_n^0 determines a deformation of M_n into an interior part of itself, and $(K_n^0 \cdot H_n)$ is then the number of points where T operates as that infinitesimal deformation. In this case

$$P_\mu = 1 = P_\mu^{11} = P_\mu^{22}; \quad P_\mu^{12} = P_\mu^{21} = 0. \tag{36.3}$$

Hence

$$(K_n^0 \cdot K_n') = \sum(-1)^\mu \text{ trace } (B_\mu Q_{11}'^{n-\mu} B_\mu^{-1} + F_\mu Q_{22}'^{n-\mu} F_\mu^{-1}). \tag{36.4}$$

But by well known properties of matrices,

$$\text{trace } uv'u^{-1} = \text{trace } v' = \text{trace } v. \tag{36.5}$$

This together with a change of μ into $n-\mu$ gives

$$(K_n^0 \cdot K_n') = (-1)^n \sum(-1)^\mu \text{ trace } Q_\mu. \tag{36.6}$$

37. A particularly interesting case of (36.2) corresponds to the fixed points of a deformation T (infinitesimal or otherwise) of M_n into an interior part of itself. Then the actual number of signed fixed points when finite is given by

$$(K_n'^0 \cdot K_n^0) = \sum(-1)^\mu R_\mu, \tag{37.1}$$

where $K_n'^0$ corresponds to the deformation. Here R_μ is the sum of the traces of P_{11}^μ and P_{22}^μ when they both reduce to the identity, that is, the sum of their orders. This is also the number of distinct Γ's and G's in the fundamental set

for the dimension μ; hence R_μ is the μth connectivity index of M_n itself, and the sum in (37.1) is its Euler characteristic. This shows that (10.6) holds for an M_n with a boundary also, of course with the K's in place of the Γ's. Therefore

THEOREM. *For every M_n, with or without boundary, the number of signed fixed points of a deformation is the Euler characteristic.*

It is understood of course that when M_n has a boundary the deformation reduces M_n to part of itself. It is necessary to point out that instead of a deformation we may equally well consider a T reducing M_n to an interior part of itself and merely adding zero-divisors to the cycles of the manifold.

The question of the *singular points of a vector distribution* is equivalent to the determination of the fixed points of an infinitesimal deformation. The authors dealing with it, notably Brouwer, Birkhoff, and recently Hopf,* have always restricted their manifolds more than in this paper. Hopf, for example, assumes that at every vertex of the defining C_n of M_n the incident cells constitute a star with the same structure as some star embedded in an S_n. Then, whatever the point A, there exists a region containing it wherein any two points may be joined by a uniquely defined polygonal line, image of a rectilinear segment in the S_n region. Until it is actually proved, as may be done for $n=2$, 3, that *every star of cells which is an n-cell has the same structure as some star in S_n*, our manifolds must be considered as much more general, and our results as having a notably wider range. On the other hand it must be stated that all analytical manifolds are of the more restricted type, so that for various applications the restriction may actually not be important.

38. The fixed point formulas derived from the general coincidence formula may also be obtained directly and in a very simple manner. Indeed, instead of associating with the identity the transformations of §§ 30, 31, we may associate with it the identical transformation for V_n also. The corresponding cycle Γ_n^0 is on $M_n \times M_n'$ and $\overline{M}_n \times M_n'$, hence at once for type I,

$$(K_n \cdot K_n^0) = (\Gamma_n \cdot \Gamma_n^0), \tag{38.1}$$

and for type II

$$(K_n^0 \cdot K_n') = (\Gamma_n^0 \cdot \Gamma_n'). \tag{38.2}$$

* Mathematische Annalen, vol. 96 (1926), pp. 225–250. He has derived (37.1) for a vector distribution on the general M_n of the type that he considers. His seem to be the only investigations on general manifolds along the line of this paper that are to be found in the literature. In Mathematische Annalen, vol. 95 (1925), he has generalized in an interesting manner the well known topological property of the total Gaussian curvature.

Then from (10.5) the desired results follow easily.

39. **Transformation of one manifold into another.** The transformations of M_n into a new manifold M_n' are treated exactly like those of M_n into itself. First let both be without boundary and let $\overline{L}_\mu$ be the analogue of L_μ for M_n'. The transformation of M_n into M_n' will be defined by cycles Γ_n, Γ_n' on $M_n \times M_n'$ and (59.2), Tr., will hold.

The expression for $(\Gamma_n \cdot \Gamma_n')$ will be again (10.4) with the first factor, L, in each term replaced by $\overline{L}$, and the others unchanged or*

$$(39.1) \qquad (\Gamma_n \cdot \Gamma_n') = \sum(-1)^\mu \operatorname{trace} \overline{L}_\mu \beta'_{n-\mu} L_\mu^{-1} \alpha_\mu .$$

When there are boundaries, the types of transformations must again be restricted as previously. If $\overline{B}$, $\overline{F}$ correspond to $\overline{L}$ as B, F to L, we find here the same coincidence formula (35.3) as before, except that B, F are replaced by $\overline{B}$, $\overline{F}$, everything else (notably B^{-1}, F^{-1}) remaining unchanged. It does not seem necessary to write the formula explicitly.

40. **A different type of coincidence formula.** In these Transactions, vol. 25 (1923), Alexander has derived for 1-τ transformations of surfaces a formula of a different type from ours. The difference consists in the fact that in place of the transformation matrices there appear everywhere the Kronecker index matrices for the intersections of the $n-\mu$ cycles with the transforms of the μ cycles. Let us show that such formulas can be derived from those of the present paper.

Let first M_n be without boundary and introduce for T the matrix

$$(40.1) \qquad \xi_\mu = \| (T\gamma_\mu^i \cdot \gamma_{n-\mu}^j) \|$$

with a similar matrix η_μ for T'. The problem is to express the number of signed coincidences in terms of the ξ's and η's. We have

$$(40.2) \qquad (T\gamma_\mu^i \cdot \gamma_{n-\mu}^j) = \left(\sum \alpha_\mu^{ih} \gamma_\mu^h \cdot \gamma_{n-\mu}^j \right)$$

$$= \sum \alpha_\mu^{ih} (\gamma_\mu^h \cdot \gamma_{n-\mu}^j),$$

therefore

$$(40.3) \qquad \xi_\mu = \alpha_\mu L_\mu ; \quad \eta_\mu = \beta_\mu L_\mu .$$

From this and the readily verified relation

$$(40.4) \qquad L'_{n-\mu} = (-1)^{\mu(n+1)} L_\mu ,$$

* See footnote, §3.

follows in place of (10.6) the desired formula

$$(\Gamma_n \cdot \Gamma_n') = \sum(-1)^{n\mu} \operatorname{trace} \eta'_{n-\mu} L_\mu^{-1} \xi_\mu L_\mu^{-1}. \tag{40.5}$$

For the fixed points, assume the second transformation to be the identity. Then directly from (10.5) or else from (40.5) and

$$\eta_{n-\mu} = L_{n-\mu} = (-1)^{\mu(n+1)} L'_\mu, \tag{40.6}$$

we find the number of signed fixed points

$$(\Gamma_n \cdot \Gamma_n^0) = \sum(-1)^\mu \operatorname{trace} \xi_\mu L_\mu^{-1}. \tag{40.7}$$

Let $n=2$ as in Alexander's paper. The terms corresponding to $\mu=0, 2$ are the same as in (10.6), namely $\alpha_0+(-1)^n \alpha_n=\alpha_0+\alpha_2$. The term $\mu=1$ alone needs to be computed. When the fundamental set is canonical, L_1 is the well known matrix $||l_{ij}||$ $(i, j=1, 2, \cdots, 2p=R_1)$ with $l_{2i-1,\ 2i}=-l_{2i,\ 2i-1}=1$ and all other terms zero. Let $\gamma_1, \cdots, \gamma_{2p}$ be the retrosections of the Riemann surface. If we recall that for any two cycles γ, δ on the surface, $(\gamma \cdot \delta)=-(\delta \cdot \gamma)$, and observe that $L_1'=-L_1$, we find that (40.7) becomes, with $\bar{\gamma}=T\gamma$ as usual,

$$(\Gamma_2 \cdot \Gamma_2^0) = \alpha_0 + \alpha_2 + \sum(\bar{\gamma}_{2i-1} \cdot \gamma_{2i}) + (\gamma_{2i-1} \cdot \bar{\gamma}_{2i}), \tag{40.8}$$

which generalizes Alexander's formula (4), loc. cit., and except for the notation reduces to it when, as he does, we consider $1-\tau$ transformations $(\alpha_0=1, \alpha_2=\tau)$.

The treatment for manifolds with a boundary is along the same line. We find

$$\begin{aligned}(K_n \cdot K_n') = \sum(-1)^{n\mu} \operatorname{trace} (&\sigma'^{n-\mu}_{11} B_\mu^{-1} \pi^\mu_{11} B_\mu^{-1} \\ &+ \sigma'_{21} N^{-1} \pi_{12} B^{-1} + \sigma'_{12} B^{-1} \pi_{21} N^{-1} \\ &+ \sigma'_{22} N^{-1} \pi_{22} N^{-1}),\end{aligned} \tag{40.9}$$

where the sequence of dimensional indices is the same for all four terms under the sum and therefore has been written down only for the first. As for the various letters their meaning is as follows:

$$\begin{aligned} B_\mu &= \| (\Gamma^i_\mu \cdot \Delta^j_{n-\mu}) \|, \quad N_\mu = \| (G^i_\mu \cdot D^j_{n-\mu}) \|, \\ \pi^\mu_{11} &= \| (T\Gamma^j_\mu \cdot \Delta^j_{n-\mu}) \| ; \end{aligned} \tag{40.10}$$

π_{12}, π_{21}, π_{22} are as π_{11} with the pair $\Gamma\Delta$ replaced by $G\Delta$, ΓD and GD respectively, while σ_{ij} is derived from π_{ij}, by substituting for T and μ, T' and $n-\mu$ and permuting the cycles of the corresponding pair.

For the fixed points of T of first type, we take $T'=1$, and it turns out that (40.7) holds also provided that ξ and L correspond to Kronecker index matrices for the fundamental sets of M_n alone, that is, for the intersections of the Γ's and G's and their transforms. For fixed points of T' of second type we take $T=1$ and find

$$(40.11) \qquad (K_n^0 \cdot K_n') = (-1)^n \sum (-1)^\mu \operatorname{trace} (B_\mu^{-1}\sigma_{11}^{\mu} + N_\mu^{-1}\sigma_{22}^{\mu}).$$

This completes the derivation of the formulas of Alexander's type from those of this paper. The type to be used in any particular case will depend upon the available data on the transformations.

PRINCETON UNIVERSITY,
PRINCETON, N. J.

ON SETS OF FUNCTIONS OF A GENERAL VARIABLE*

BY

LLOYD L. DINES

A theory of functions of a general variable is due to E. H. Moore.† By a general variable is meant a variable of which the range is a class of elements

$$\mathfrak{Q} \equiv [q]$$

entirely unconditioned. Particular instances of the theory may be obtained by specializing the class $\mathfrak{Q}$. For example the class $\mathfrak{Q}$ may consist of a finite number of elements, a denumerable infinitude of elements, or a continuous infinitude of elements. The elements themselves may be numbers, real, complex, or hypercomplex; or they may be without numerical character.

A real (single-valued) function μ on a general range $\mathfrak{Q}$ is a correspondence between the elements of $\mathfrak{Q}$ and a class of real numbers, such that for every element q of $\mathfrak{Q}$ there is a definite corresponding real number, notationally $\mu(q)$.

A property of such a function which is in no way dependent for its definition on any special character which the range $\mathfrak{Q}$ may have in special instances is said to be a *property of general reference*. For example, a function may be: (a) everywhere zero on $\mathfrak{Q}$, or (b) everywhere positive on $\mathfrak{Q}$, or (c) everywhere negative on $\mathfrak{Q}$, or (d) somewhere positive and nowhere negative on $\mathfrak{Q}$, or (e) somewhere negative and nowhere positive on $\mathfrak{Q}$. We shall use the following symbolic statements to indicate that a function μ has these properties respectively:

(a) $\mu = 0\ (\mathfrak{Q})$, (d) $\mu \geqq' 0\ (\mathfrak{Q})$,

(b) $\mu > 0\ (\mathfrak{Q})$, (e) $\mu \leqq' 0\ (\mathfrak{Q})$.

(c) $\mu < 0\ (\mathfrak{Q})$,

In the present paper we shall be particularly interested in functions which have the property (d) or the property (e). A function which has either of these properties will be said to be *M-definite*. In other words an M-definite function is one which is not identically zero and does not change sign on $\mathfrak{Q}$.

* Presented to the Society, San Francisco Section, June 12, 1926; received by the editors in July, 1926.

† The theory is called by Professor Moore "General Analysis," and is developed in his New Haven Mathematical Colloquium Lectures, New Haven, 1910.

Properties of general reference may pertain to a *set* of real functions

$$\mu_1, \mu_2, \cdots, \mu_m,$$

each on the general range $\mathfrak{Q}$.* For example, if there exists a set of real constants $c_1, c_2, \cdots, c_m$ (not all zero) such that

$$\sum_{j=1}^{m} c_j\mu_j = 0 \quad (\mathfrak{Q}), \tag{1}$$

the set of functions is said to be linearly dependent; otherwise it is linearly independent. Other properties of general reference for a set of functions are obtained by replacing the sign $=$ in (1) by the signs used in (b)-(e). For the special instance in which $\mathfrak{Q}$ consists of a finite number of elements, these properties have been considered by the author.†

The object of the present paper is to study the condition

$$\sum_{j=1}^{m} c_j\mu_j \geqq' 0 \quad (\mathfrak{Q}), \tag{2}$$

that is, the condition that a given set of functions on a general range admit an M-definite linear combination.

The central feature of the theory is a certain integral-valued function of the set of functions $\{\mu_j\}$ which we shall call the *M-rank* of the set. In terms of it may be stated a necessary and sufficient condition that (2) admit a solution $(c_1, c_2, \cdots, c_m)$ and the maximum number of c's that may be zero in such a solution. These results are stated in §4, the earlier sections being preparatory to the definition of M-rank.

1. **Reduction and composition of a general range $\mathfrak{Q}$ relative to a function on that range.** We consider a class $\mathfrak{Q}$ of elements q, notationally

$$\mathfrak{Q} \equiv [q].$$

Let μ be any real single-valued function on $\mathfrak{Q}$. We shall have occasion to consider three subclasses of $\mathfrak{Q}$, relative to μ, defined as follows:

$$\mathfrak{Q}_P{}^{(\mu)} \equiv [\text{all } q \text{ such that } \mu(q) > 0] \equiv [p^{(\mu)}],$$
$$\mathfrak{Q}_N{}^{(\mu)} \equiv [\text{all } q \text{ such that } \mu(q) < 0] \equiv [n^{(\mu)}],$$
$$\mathfrak{Q}_Z{}^{(\mu)} \equiv [\text{all } q \text{ such that } \mu(q) = 0] \equiv [z^{(\mu)}].$$

* It may be noted, however, that a property of such a *set* of functions can be considered as a property of a *single* function μ' on a composite range $\mathfrak{Q}'$, the elements q' of the range $\mathfrak{Q}'$ being bipartite elements of the form $q' = (q, j)$, the first part q having the range $\mathfrak{Q}$ and the second part j having the finite range consisting of the numbers $1, 2, \cdots, m$.

† Annals of Mathematics, (2), vols. 20, 27, and 28.

From these subclasses we form a certain composite range relative to the function μ,

$$\mathfrak{Q}^{(\mu)} \equiv \mathfrak{Q}_P{}^{(\mu)}\mathfrak{Q}_N{}^{(\mu)} + \mathfrak{Q}_Z{}^{(\mu)},$$

consisting of the logical sum of the two classes

$$\mathfrak{Q}_P{}^{(\mu)}\mathfrak{Q}_N{}^{(\mu)} \equiv [p^{(\mu)}n^{(\mu)}] \text{ and } \mathfrak{Q}_Z{}^{(\mu)} \equiv [z^{(\mu)}].$$

The elements of the new class

$$\mathfrak{Q}^{(\mu)} \equiv [q^{(\mu)}]$$

will therefore be of two kinds: (1) bipartite elements $p^{(\mu)}n^{(\mu)}$ of which the first part $p^{(\mu)}$ ranges over $\mathfrak{Q}_P{}^{(\mu)}$ and the second part $n^{(\mu)}$ ranges over $\mathfrak{Q}_N{}^{(\mu)}$ independently; and (2) unipartite elements $z^{(\mu)}$ ranging over $\mathfrak{Q}_Z{}^{(\mu)}$.

The process here indicated may evidently be repeated. If σ is a real single-valued function on the new range $\mathfrak{Q}^{(\mu)}$, it determines three subclasses of the range, which may be denoted by $\mathfrak{Q}_P{}^{(\mu\sigma)}$, $\mathfrak{Q}_N{}^{(\mu\sigma)}$, and $\mathfrak{Q}_Z{}^{(\mu\sigma)}$; and from these may be formed the composite class

$$\mathfrak{Q}^{(\mu\sigma)} \equiv \mathfrak{Q}_P{}^{(\mu\sigma)}\mathfrak{Q}_N{}^{(\mu\sigma)} + \mathfrak{Q}_Z{}^{(\mu\sigma)}.$$

The process may be repeated indefinitely, provided at each stage a reducing function is available.

It will be noted that if at any stage the reducing function is everywhere positive or everywhere negative, the new composite range will be a null class; while if the reducing function is identically zero, the new composite range will be identical with the old range.

2. **Reduced outer multiplication.** The reducing function μ determines with any second function ν on the range $\mathfrak{Q}$, a real single-valued function on the composite range $\mathfrak{Q}^{(\mu)}$ which we will call their *reduced outer product*, and denote by $((\mu\nu))$. It is defined as follows:*

$$((\mu\nu)) \equiv \begin{cases} \mu(p)\nu(n) - \nu(p)\mu(n) & \text{for } pn \text{ on } \mathfrak{Q}_P{}^{(\mu)}\mathfrak{Q}_N{}^{(\mu)}, \\ \nu(z) & \text{for } z \text{ on } \mathfrak{Q}_Z{}^{(\mu)}. \end{cases}$$

This multiplication is not commutative. Its most obvious property is that $((\mu\mu))=0$ on $\mathfrak{Q}^{(\mu)}$. Other properties are developed in the next section.

* The outer product of two functions $f(x)$ and $g(x)$, where x is a real variable on a closed interval, has been defined as $f(x)g(y)-g(x)f(y)$. See Kowalewski's *Funktionenräume*, Wiener Sitzungsberichte, vol. 120.

3. **Reduction of a set of functions.** Consider a set of real, single-valued functions

$$\{\mu_j\} \qquad \mu_1, \mu_2, \cdots, \mu_m,$$

on a general range $\mathfrak{Q}$.

Relative to any one of the functions, say μ_k, we may determine a second set of m functions

$$\{\mu_j^{(k)}\} \qquad ((\mu_k\mu_1)), ((\mu_k\mu_2)), \cdots, ((\mu_k\mu_m)),$$

each of which is the reduced outer product of the corresponding function in the given set by μ_k. This new set of functions on the composite range $\mathfrak{Q}^{(\mu_k)}$ will be called a reduced set, or more explicitly, the reduction of the set $\{\mu_j\}$ with respect to μ_k. It has the notable property that its kth constituent is identically zero. The usefulness of this type of reduction lies in the following lemma:

If the reducing function μ_k is not M-definite, then the set $\{\mu_j\}$ admits an M-definite linear combination if and only if the same is true of the reduced set $\{\mu_j^{(k)}\}$. More explicitly, every set of constants $c_1, c_2, \cdots, c_m$, satisfying the condition

$$\sum_{j=1}^{m} c_j\mu_j \geqq' 0 \quad (\mathfrak{Q}), \tag{2}$$

will also satisfy the reduced condition

$$\sum_{j=1}^{m} c_j((\mu_k\mu_j)) \geqq' 0 \quad (\mathfrak{Q}^{(\mu_k)}); \tag{3}$$

and conversely, every solution $c_1, c_2, \cdots, c_m$ of (3) yields a solution of (2) if the constant c_k, which is arbitrary in a solution of (3), be suitably chosen.

We note first that if $\mu_k = 0(\mathfrak{Q})$, the proposition is true though trivial, since in that case the two sets $\{\mu_j\}$ and $\{\mu_j^{(k)}\}$ are identically the same. We may then assume in the proof that the function μ_k changes sign on $\mathfrak{Q}$.

Suppose that the condition (2) has a solution $\bar{c}_1, \bar{c}_2, \cdots, \bar{c}_m$.

Taking account of the three subclasses of the range $\mathfrak{Q}$ relative to the function μ_k, we obtain from this hypothesis the three statements

$$\sum_{j=1}^{m} \bar{c}_j\mu_j(p) \geqq 0 \qquad (p \text{ on } \mathfrak{Q}_P^{(\mu_k)}), \tag{4}$$

$$\sum_{j=1}^{m} \bar{c}_j\mu_j(n) \geqq 0 \qquad (n \text{ on } \mathfrak{Q}_N^{(\mu_k)}), \tag{5}$$

$$\sum_{j=1}^{m} \bar{c}_j \mu_j(z) \geqq 0 \qquad (z \text{ on } \mathfrak{Q}_z^{(\mu_k)}), \tag{6}$$

with the understanding that the sign $\geqq$ has the significance of $\geqq'$ in at least one of the three statements.

Multiplying (4) by $-\mu_k(n)$ and (5) by $\mu_k(p)$ and adding the results, we have

$$\sum_{j=1}^{m} \bar{c}_j [\mu_k(p)\mu_j(n) - \mu_j(p)\mu_k(n)] \geqq 0 \qquad (pn \text{ on } \mathfrak{Q}_P^{(\mu_k)}\mathfrak{Q}_N^{(\mu_k)}).$$

This together with (6) may be written

$$\sum_{j=1}^{m} \bar{c}_j((\mu_k\mu_j)) \geqq' 0 \quad (\mathfrak{Q}^{(\mu_k)}), \tag{7}$$

which proves the first part of the proposition.

Suppose conversely that the condition (3) has a solution $\bar{c}_1, \bar{c}_2, \cdots, \bar{c}_m$, as expressed by (7).

We note first that the constant $\bar{c}_k$ is arbitrary, since its coefficient is identically zero. We may therefore omit the term corresponding to $j=k$ from the summation, indicating the omission by an apostrophe', and write (7) in the two statements

$$\sum_{j=1}^{m}{}' \bar{c}_j [\mu_k(p)\mu_j(n) - \mu_j(p)\mu_k(n)] \geqq 0 \qquad (pn \text{ on } \mathfrak{Q}_P^{(\mu_k)}\mathfrak{Q}_N^{(\mu_k)}), \tag{8}$$

$$\sum_{j=1}^{m}{}' \bar{c}_j \mu_j(z) \geqq 0 \qquad (z \text{ on } \mathfrak{Q}_z^{(\mu_k)}), \tag{9}$$

one of the signs $\geqq$ having the significance of $\geqq'$.

Since $-\mu_k(p)\mu_k(n)$ is positive, we may obtain from (8) an equivalent statement

$$\sum_{j=1}^{m}{}' \bar{c}_j \frac{\mu_j(p)}{\mu_k(p)} \geqq \sum_{j=1}^{m}{}' \bar{c}_j \frac{\mu_j(n)}{\mu_k(n)} \qquad (\mathfrak{Q}_P^{(\mu_k)}\mathfrak{Q}_N^{(\mu_k)}). \tag{10}$$

Now the values on the left side of (10) must have a greatest lower bound, and those on the right a least upper bound, which bounds may or may not coincide. In any case we may choose the arbitrary $\bar{c}_k$ so that

$$\sum_{j=1}^{m}{}' \bar{c}_j \frac{\mu_j(p)}{\mu_k(p)} \geqq -\bar{c}_k \geqq \sum_{j=1}^{m}{}' \bar{c}_j \frac{\mu_j(n)}{\mu_k(n)}.$$

And from this double relation we obtain

$$\sum_{j=1}^{m}{}' c_j\mu_j(p) + c_k\mu_k(p) \geqq 0, \qquad \sum_{j=1}^{m}{}' c_j\mu_j(n) + c_k\mu_k(n) \geqq 0,$$

which together with (9) may be written

$$\sum_{j=1}^{m} c_j\mu_j \geqq' 0 \quad (\Omega).$$

This completes the proof of the lemma.

The process of reduction may be repeated. As reduction of the set $\{\mu_j\}$ with respect to μ_k yields the set $\{\mu_j^{(k)}\}$, so reduction of this latter set with respect to one of its constituents $\mu_l^{(k)}$ yields a set which we shall denote by $\{\mu_j^{(kl)}\}$.

In general, we define the set

$$\{\mu_j^{(k_1k_2\cdots k_s)}\} \tag{11}$$

as the reduction of the set

$$\{\mu_j^{(k_1k_2\cdots k_{s-1})}\}$$

with respect to the function $\mu_{k_s}^{(k_1k_2\ldots k_{s-1})}$.

The set (11) will be called an sth reduction of the set $\{\mu_j\}$. Clearly there are many sth reductions, depending on the choice of the sequence $k_1, k_2, \cdots, k_s$. If the integers in this sequence are distinct, s constituent functions of the sth reduction are identically zero.

4. **The M-rank of a set of functions.** We recall that a function μ is said to be M-definite if either of the conditions

$$\mu \geqq' 0 \ (\Omega) \text{ or } \mu \leqq' 0 \ (\Omega)$$

is satisfied.

A set of m functions $\{\mu_j\}$ is said to be of M-rank r if at least one of its $(m-r)$th reductions contains an M-definite constituent function while no one of its $(m-r-1)$th reductions contains such a constituent function. If the given set contains an M-definite function the set is of M-rank m. If neither it nor any of its reductions contains such a function it is of M-rank zero.

THEOREM. *A necessary and sufficient condition that the set of m functions admit an M-definite linear combination is that its M-rank be greater than zero.*

If the M-rank is $r(0<r<m)$, then there is a subset of $m-r+1$ of the functions which admits an M-definite linear combination, but there is no subset of $m-r$ functions for which this is true.

First, if the given set $\{\mu_j\}$ admits an M-definite linear combination, its M-rank is greater than zero. For otherwise the $(m-1)$th reduction $\{\mu_j^{(1,2,\cdots,m-1)}\}$ would contain no M-definite function, while the corresponding reduced condition

$$\sum_{j=1}^{m} c_j \mu_j^{(1,2,\cdots,m-1)} \geqq' 0 \qquad (\mathfrak{Q}^{(1,2,\cdots,m-1)})$$

must admit a solution $c_1, c_2, \cdots, c_m$ by the lemma of §3. These two requirements are incompatible since all functions of the $(m-1)$th reduction are zero except one.

Conversely, suppose the M-rank of the given set is $r(>0)$. Then there is an $(m-r)$th reduction of the set which contains an M-definite function. Suppose, for simplicity of notation, it is $\{\mu_j^{(1,2,\cdots,m-r)}\}$, and suppose the $(m-r+1)$th function of this set is M-definite. Then the condition

$$\sum_{j=1}^{m} c_j \mu_j^{(1,2,\cdots,m-r)} \geqq' 0 \qquad (\mathfrak{Q}^{(1,2,\cdots,m-r)})$$

admits a solution $c_1, c_2, \cdots, c_m$, in which

$c_j=0$ for $j>m-r+1$,

c_j is arbitrary for $j<m-r+1$,

$c_{m-r+1}=+1$ or -1 according as $\mu_{m-r+1}^{(1,2,\cdots,m-r)}$ is positive or negative.

Hence by repeated application of the lemma of §3, we find that the condition

$$\sum_{j=1}^{m} c_j \mu_j \geqq' 0 \qquad (\mathfrak{Q})$$

admits a solution in which $c_j=0(j>m-r+1)$, suitable values being assigned to $c_j(j<m-r+1)$. The given set therefore admits an M-definite linear combination; indeed a subset of $m-r+1$ of them has this property.

It remains to be proved that when the M-rank is r, no subset of $m-r$ of the given functions admits an M-definite linear combination. Suppose for definiteness that the subset

(12) $\qquad \mu_1, \mu_2, \cdots, \mu_{m-r}$

did this property. Then by the first proposition of the theorem (already established) the M-rank of this subset must be greater than zero, call it r'. That means that some $(m-r-r')$th reduction of the subset (12) would contain an M-definite constituent function. The corresponding $(m-r-r')$th reduction of the original set of m functions would contain the same M-definite function, and hence the M-rank of the set would be $r+r'$, contrary to our assumption that it was r.

This completes the proof of the theorem.

UNIVERSITY OF SASKATCHEWAN,
SASKATOON, CANADA